Exhibit D

The Weight of Scientific Evidence in Policy and Law

Sheldon Krimsky, PhD

The term "weight of evidence" (WOE) appears in regulatory rules and decisions. However, there has been little discussion about the meaning, variations of use, and epistemic significance of WOE for setting health and safety standards.

This article gives an overview of the role of WOE in regulatory science, discusses alternative views about the methodology underlying the concept, and places WOE in the context of the Supreme Court's decision in *Daubert v Merrell Dow Pharmaceuticals, Inc* (1993). I argue that whereas the WOE approach to evaluating scientific evidence is gaining favor among regulators, its applications in judicial processes may be in conflict with some interpretations of how the Daubert criteria for judging reliable evidence should be applied. (*Am J Public Health.* 2005;95:S129–S136. doi: 10.2105/AJPH.2004.044727)

In the narratives describing the historical development of natural science, nothing captures the drama of discovery as effectively as the "crucial experiment" (an *experimentum crucis*). For it is such an experiment, according to most historical accounts, that finally resolves competing explanations and/or theories, bringing to a close contested schools of thought. For example, history of science texts tell us that it was a crucial experiment that put to rest the theory of spontaneous generation in favor of the germ theory of disease and that launched a critical blow to the Phlogiston theory of combustion. Also widely acclaimed as a crucial experiment in the early part of the 20th century were the measurements made by British physicists, among them Sir Arthur Eddington, of the sun's rays during a solar eclipse. From their measurements they concluded that light bends in a gravitational field, which provided evidence in support of Einstein's over Newton's theory of light.[1] Such experiments have gained iconic status in the history of science.

But there is a significant and lively debate among philosophers and historians on whether it is meaningful to talk about "crucial experiments" in science. Sir Karl Popper believed that crucial experiments could play a role in falsifying scientific theories ("It should be noted that I mean by a crucial experiment one that is designed to refute a theory (if possible) and more especially one which is designed to bring about a decision between two competing theories by refuting (at least) one of them—without of course, proving the other.")[2] In contrast, Pierre Duhem and Thomas Kuhn were leading voices against the view that scientists are influenced by "crucial experiments" in deciding between competing paradigms.

Notwithstanding this debate, I believe there are influential or determinative experiments that crystallize a new scientific consensus, particularly in fields like physics, chemistry, and engineering.

However, it is very rare to find determinative experiments in environmental health sciences. A single, well-constructed experiment almost never resolves a critical issue on the cause of a disease, especially but not exclusively, diseases resulting from exposure to toxic substances. Rothman provides an example where the etiology of "toxic shock syndrome" was resolved through a crucial experiment.[3] As long as we do not permit controlled experiments where we would intentionally harm a human subject, when there are no possible benefits to them, for the mere sake of scientific inquiry, no single experiment can provide decisive data on the effects of a foreign substance on a human group. In so far as we depend on a number of experiments, some with greater statistical or explanatory power than others and information from diverse forms of evidence, we need to have some way of aggregating or weighing the results across different modalities of evidence.

The term "weight of evidence" (WOE) is used to characterize a process or method in which all scientific evidence that is relevant to the status of a causal hypothesis is taken into account. In criminal law, juries are given the responsibility to decide the WOE in regards to guilt or innocence. Judges weigh the evidence of legal precedent in justifying their rulings. Clinicians use a form of WOE in making diagnoses, and judges may defer to it when they offer opinions on the reliability of evidence. In the policy sectors of government, regulatory agencies or risk analysis panels use WOE to assess the total value of the scientific evidence that a substance may be dangerous to human health. Sometimes the term is used as if there were some algorithm or rational decision process by which the "weighing of evidence" is accomplished. Other times, the term WOE refers to nothing more than a subjective assessment on the part of a reviewer who takes relevant data from a given body of published research into consideration in order to ascertain whether a hypothesis is more likely to be true than false.

A distinction has been made between WOE and "strength of evidence" (SOE).[4] The latter is associated with the gravitas and relevance of information related to a specific indicator, such as the number of tumors produced in animals. In contrast, WOE includes all varieties of evidence, positive and negative, mechanistic and nonmechanistic, in vivo and in vitro, as well as human and animal studies. In risk assessment, the trend has been to widen the lens of relevant empirical and theoretical evidence, thus moving from approaches that utilize "strength of evidence" to those that utilize WOE. In this article I shall speak exclusively of WOE and assume that it encompasses the use of strength of evidence.

The WOE approach has been introduced into ecological risk assessment since the early 1990s in response to the need for better risk analyses of Superfund sites and impacted natural ecosystems.[5,6] One consensus report on WOE defined it as "the process by which multiple measurement endpoints are related to an assessment endpoint to evaluate whether a significant risk of harm is posed to

the environment."[7] In his widely cited book *Ecological Risk Assessment*, Suter notes the significance of WOE in evaluating different classes of evidence generated by alternative ecological models. He wrote, "the separate lines of evidence must be evaluated, organized in some coherent fashion, and explained to the risk manager so that a weight of evidence evaluation can be made."[8]

A number of benefits have been attributed to a WOE framework in regulatory decisions. Walker[9] cites three objectives of a WOE analysis: (1) it provides a "clear and transparent framework" for evaluating the evidence in a risk determination; (2) it offers regulatory agencies a consistent and standardized approach to evaluating toxic substances; and (3) it helps to identify the discretionary assumptions in risk determinations from experts. However, in selecting a WOE approach a certain number of nontestable *a priori* assumptions must be adopted, which may narrow the scope of scientific opinion and consensus on how different modalities of evidence should be aggregated, thereby failing to meet Walker's objective.

I begin with the observation that there is virtually no discussion in the scientific literature of the epistemic meaning of WOE. We cannot tell whether it is used as a methodology, a heuristic, a ranking system, or simply a subjective process of setting a causal threshold for cumulative indirect evidence. In the spirit of these questions, this article will do the following: (1) discuss the problem of aggregating different forms of evidence; (2) review uses of WOE in health science publications; (3) examine some applications of WOE by federal agencies; and (4) discuss how WOE enters judicial proceedings, particularly in the context of the admissibility of expert witnesses.

In this discussion, I shall argue that the concept of WOE, as it is currently applied in the health sciences, largely involves a qualitative approach to rating and assessing the aggregation of different forms of scientific evidence in relationship to a causal hypothesis. Currently, qualitative or quantitative frameworks that guide the use of a WOE method are more or less *a priori* heuristics that adopt certain norms about the status and relevance of alternative types of information, but their application largely depends on the tacit expertise of scientific evaluators. Moreover, no canonical frameworks for weighing scientific evidence have emerged. When experts use the term WOE in publications or in the courtroom, they are almost always referring to the outcome of a process in which scientists, working as individuals or in groups, examine a body of relevant scientific studies on the relationship between a compound and a disease outcome. These scientists, operating within an accepted framework, apply their tacit knowledge of a field to reach a "yea," "nay," or "probabilistic conclusion" about the relationship between the compound and a disease outcome. Most applications of WOE in support of public policy that are cited in the literature seem to (by inference or lack of specification) use a process methodology that is low on transparency and high on subjectivity.

MODALITIES OF EVIDENTIARY SUPPORT

If the modality of evidence considered for evaluating the human health effects of a chemical compound was of one type, let us say epidemiological studies, then the WOE might refer to how many studies support the hypothesis about health risks, what the individual power of a study is, or what the combined power of all the studies are in a meta-analysis. But each modality of evidentiary support is limited. For example, some scientists argue that epidemiological studies cannot demonstrate causation or mechanism, but only association.[10] Controlled animal studies do not yield direct information about people. Comparison of chemical structure between suspected and known toxins (known as structure activity analysis) does not provide information on how the chemicals function in a live organism. The term WOE has come to mean not only a determination of the statistical and explanatory power of any individual study (or the combined power of all the studies) but the extent to which different types of studies converge on the hypothesis. The WOE approach has become likened to "triangulation," namely, approaching the target question from many directions. Where no single epistemic modality (by which I mean a specific method or technique for acquiring information) can yield the definitive answer to an environmental health question, we refer to multiple epistemic modalities. The problem is: how does the evidence from these modalities add up? Does the accumulated data from several epistemic modalities mitigate against the insufficiency or shortcomings of evidence from a single epistemic modality?

A similar problem is presented in decision analysis. Multiattribute Utility Theory applies to cases where there are different dimensions of value associated with outcomes that, on the face of it, are not reducible to a common metric.[11] Thus, a decision to build a dam will have both positive and negative impacts of a social and ecological variety. These attributes are incommensurable, such as the additional hydropower gained by the dam and the loss of fish spawning in the river. In Multiattribute Utility Theory, a decision analyst develops a ranking and a utility function for the attributes and then undertakes an empirical investigation to determine the actual value of those attributes (how many fish will be lost and how much energy will be produced). Thus, the final outcome of applying Multiattribute Utility Theory is the aggregation of incommensurable variables through the adoption of a numerical schema.

For evaluating the human health effects of a chemical agent, there are different modalities of evidence, including human epidemiology, wildlife studies, experimental laboratory animal studies with rodents, primate studies, *in vitro* cell studies, and chemical structure activity analysis. Each type of study may provide some evidence, but each has its limitations. Human epidemiology may be valued highly for its relevance but less so for its scientific power, especially if the findings are unrelated to a postulated or known biological mechanism. Experimental animal studies may be dependable for the mechanistic knowledge they offer but questionable for their relevance to human cases.

If a chemical were known to be one of the causal agents responsible for a human disease, then we would expect a series of evidentiary pathways to converge on that conclusion. The chemical might manifest genotoxic or gross chromosomal effects in human cells studied *in vitro*. Or the chemical might be associated with wildlife abnormalities. But not all of the

evidence may be consistent with the result. It is possible that the chemical may be harmless to certain species and yet cause disease in others. Nevertheless, we gain confidence when one epistemic modality (rodent studies) is consistent with the results of other epistemic modalities (epidemiological studies) that make up the architectonic of evidence.

When we do not *know* whether a chemical causes a human disease but have the type of circumstantial evidence we would expect to acquire if the substance were known to cause the disease, then, building on a coherence theory of truth, the weight of the circumstantial or related evidence elevates our confidence in the hypothesis connecting the substance to the disease.

But how can we aggregate the evidence from a variety of modalities in a WOE approach, when no single study is definitive, and we cannot justifiably reach a conclusion from the limited evidence that a specific compound is likely the cause of human illness? Aggregating evidence across different epistemic modalities is like adding incommensurables. It can only be done if *a priori* constructs provide a basis for developing a common metric. More evidence, albeit inconclusive, may mean you are closer to demonstrating causality, but you cannot know by how much. And sometimes, different modalities of evidence do not converge on a single hypothesis and may even be inconsistent.

USES OF THE TERM WOE IN HEALTH SCIENCES

Usually WOE methods are applied when no single study and no individual modality of evidence (e.g., animal studies, human studies, *in vitro,* etc.) is conclusive in demonstrating a cause-effect relationship. Other times it may be used even when there is a solid epidemiological study showing a large increased risk from the exposure to some substance in order to build a stronger argument for regulation. Alternatively, WOE has been introduced to assess the "strengths and weaknesses of various measurements, and of the nature of uncertainty associated with each of them."[12] However, while the term is applied quite liberally in the regulatory literature, the methodology behind it is rarely explicated. We might be told that the decision to regulate was decided on the WOE rather than a crucial study demonstrating causality. Without an explication of how evidence is "weighed" or "weighted," the claim WOE seems to be coming out of a "black box" of scientific judgment.[13] One article that uses the term WOE in its title does not refer to the term elsewhere in the text.[14] Other articles assign scaling factors or qualitative terms to the evidentiary attributes.

A report issued by the US Agency for Toxic Substances and Disease Registry (ATSDR) of the Department of Health and Human Services stated that a necessary and reasonable alternative to causal determinations when establishing policy "may be a critical assessment of the overall "weight of evidence" of available science to serve as a surrogate of 'causality.'" The implication is that when causality is out of reach, we must use a surrogate called WOE. The ATSDR states: "'The weight-of-evidence' approach is an assessment method that includes reviewing site-specific doses, epidemiologic studies, and chemical-specific toxicity data to evaluate exposures and potential health effects in a community."[15]

In law, when direct material evidence of a crime or direct eyewitness testimony is not available, the term "circumstantial evidence" is used. This type of evidence comes in "bundles" and eventually must be "weighed" by the jury in its role of determining guilt or innocence. Each piece of the "bundle" of circumstantial evidence is insufficient to make a case. It is the entire "bundle" that convinces the jury. The concept of "circumstantial evidence" has a counterpart in environmental health.

The ATSDR uses the metaphor of the microscope as the rationale for applying the WOE approach to examining the human effects of polychlorinated biphenyls, by aggregating the results of disparate studies.

"Each of the studies, whether an epidemiologic study, a laboratory study, or the findings of wildlife biologists, could be compared to the lens of a microscope. Like the lens of a microscope, they can vary in terms of their resolving power and quality. They are also focused on different populations at different points in time . . . . Despite the limits and weaknesses of individual pieces of research, the collective weight of evidence indicates that certain polychlorinated biphenyl/dioxin-like compounds found in fish in the Great Lakes-St. Lawrence basin and elsewhere can cause neurobehavioral deficits."[16]

The concept of WOE is used widely but rarely explicated in the scientific and policy literature. Menzie et al.[17] state that, "although the term 'weight-of-evidence' is used frequently in ecological risk assessment, there is no consensus on its definition or how it should be applied." Often when WOE is cited, it is assumed that readers know what it means. Sometimes it is used to signify that evidence must reach a certain critical threshold before it can support regulation. Other times it refers to a process that examines both positive and negative studies and determines by the number and strength of the studies whether a causal relationship can be inferred. As regulatory bodies and scientific review panels depend increasingly on WOE methods, questions surrounding their use will inevitably enter litigation either in torts or contested regulations, where the elusive methodology behind WOE is ripe for Daubert challenges. Therefore, it is important to understand how WOE is being interpreted and what, if any, criteria are implicit or explicit in its application.

After an extensive review of the appearance of WOE in public health studies and regulatory documents, I have uncovered what I believe are four general uses of the term.

Intensive Literature Review

This interpretation of WOE takes the form of an intensive literature review, including some qualitative discussion of the studies, without assigning any weights to the studies. In the words of one medical group, "the more inclusive method of literature review involves assessing the 'weight of evidence' . . . the importance of the findings from each piece of research should be judged: this is termed 'Signal.' This is then balanced by the strength of the evidence or design weaknesses (termed 'Noise')."[18] Those who use the term WOE in this context assume that the reviewers have applied their expertise in interpreting both the quantity (number of positive studies) and the quality (statistical power) of the evidence without any explicit reference to a methodology. Readers may justifiably assume that the

reviewers are basing their interpretation of the aggregate value of the selected studies on their years of experience and tacit knowledge, rather than a fully developed analytical framework.[19]

Seat-of-the Pants Qualitative Assessment

According to this view, WOE is a vague term that scientists use when they apply implicit, qualitative, and/or subjective criteria to evaluate a body of evidence. Experts cite the general grounds for their opinion, but no specific parameters or methodologies are given for how the evidence is weighed. Thus, one might see general statements such as: A decision was made based on WOE standards, such as number of studies, strength of association, breadth and consistency of evidence, correlational power, and biological plausibility. A number of papers use the term WOE in the title without explaining a methodology or process that is used to do the weighting.

Sometimes the application of WOE involves a taxonomic presentation of studies. An example can be found in a 2001 study of "disinfection by-products."[20] These are the potential human hazards of chlorination. The authors created a table of evidence, which listed the summary data of studies for each adverse reproductive effect focusing on sample size, exposure assessment, relative risk, and odds ratios. They describe as the goal of the paper "to view the totality of the evidence in order to judge the overall weight of evidence concerning 'disinfection by-products' and reproductive and developmental effects."[21] After commenting on the categories listed in their taxonomy (odds ratios, uncertainties, and statistical significance), the authors conclude that the weight of evidence shows that low birth weight is not associated with "disinfection by-products" exposure. But the outcome they reach is not logically or rigorously derived by a methodology. The justification for the use of WOE could be enhanced if criteria for weighing the evidence were established at the outset.

Aggregating Diverse Evidentiary Modalities

In this particular use of WOE, an effort is made to aggregate the evidence through some combination of qualitative and/or quantitative techniques. For example, ATSDR incorporates an assessment method that includes reviewing site-specific doses, epidemiological studies, and toxicity data. A dose level injurious to humans is found from different types of research protocols.

The World Health Organization's Global Assessment of Endocrine Disrupting Chemicals uses "overall strength of evidence" as a qualitative evaluation of the outcome of concern and an exposure to a substance—assessing the strength of association as weak, moderate, or strong based on the qualitative values of each of five evaluation factors.[22]

Calabrese et al.[23] have proposed a quantitative ranking scheme to evaluate the endocrine effects of chemicals on human health. In their scheme endocrine disruption is considered a multistage process, where they assume the probability of achieving the end result, namely a clinical endocrine pathology, rises as one progresses through the process. The authors identify five levels of evidence that correspond with the stages of the multistage process, level 1 being the weakest and level 5 the strongest. Then they introduce a point system based on a geometric progression ($a + ar + ar^2 + ar^3 + ar^4$), which is normalized to 10 points when stage 5 is reached. Stages 1-4 are weighted as 0.6, 1.3, 2.5, and 5.0, respectively. The causal chain is neither linear nor deterministic. Stage 3 will not always reach stage 5, but only does so at a certain probability. Therefore, by attaching a weight to each stage, one is essentially assigning probability estimates to the evidence. Thus, these weights represent the probabilities that the specific stage will proceed to the next stage.

In theory it is possible to come up with weighting factors that are empirically verifiable. Let us suppose we are trying to determine whether a chemical is a human endocrine disruptor (that it will adversely affect the human endocrine system) and that there are five stages in the causal chain. If we had evidence that the chemical induced stage 5 effects, then we can declare the substance a human endocrine disruptor. Let us assume we have evidence the chemical induced a stage 3 effect. If we had a toxicological database with thousands of entries that allowed us to calculate the percentage of those chemicals that induced a stage 3 effect and the frequency among those that also induced a stage 5 effect, we would have an empirically based system to develop weighting factors.

However, there is no generally accepted rationale for such *a priori* weightings within a discipline. And if there were an accepted framework of weightings, the selection would be premised on achieving consistency among expert evaluators rather than on some consensus about causality.

WOE in Hypothesis Testing

Sometimes the term WOE refers to a methodology used for selecting between two competing hypotheses. In this context, authors refer to WOE in the quantitative evaluation of a hypothesis relative to the null hypothesis, based on *a priori* evidence.[24] It is common to find Bayesian methods of analysis used, where the probability of a hypothesis is based on current evidence and prior probabilities. This use of WOE is discussed in a published report that examines whether a DNA profile of a suspect is unique in the population.[25] A suspect's DNA is compared to the DNA found at the crime scene. The comparison is presented in the form of a probability estimate that the suspect's DNA and the DNA found at the crime scene are a perfect match. The weight of evidence is synonymous with the probability estimate.[26]

THE FEDERAL AGENCY USE OF WOE

US Federal agencies, as well as international agencies like the International Joint Commission,[27] have begun to incorporate WOE in both their internal risk assessment analysis and in their advisory processes where they engage with external experts. The approaches taken are usually qualitative and avoid compressing all of the data to some WOE numerical value. The ATSDR uses a WOE approach to evaluate the synergistic effects of chemical mixtures.[28] The ATSDR describes the objectives of and factors to consider in a WOE analysis in its *Public Health Assessment Guidance Manual*, without providing any details on how evidence is actually "weighed" or scaled.[29]

"A weight-of-evidence analysis involves the balanced review and integration of relevant

exposure, toxicologic, epidemiologic, medical, and health outcome data to help determine whether exposure to contaminant levels under site-specific conditions might result in harmful effects. . . . The goal of the weight-of-evidence analysis is to decide whether or not harmful effects might be possible in the exposed population by weighing the scientific evidence and by keeping site-specific doses in perspective."[30]

The Occupational Safety and Health Administration (OSHA) has incorporated WOE in its regulations. In OSHA's air contaminants standard the agency stated:

> In response to those commenters who argued that none of the studies described by OSHA presented sufficient dose-response data to be used as a basis for establishing a limit, the Agency emphasizes that it is not relying on any single study to determine that wood dust presents a significant risk of material health impairment. Instead, OSHA is making this determination on the basis of the findings in the dozens of studies reporting on the respiratory, irritant, allergic, and carcinogenic properties of wood dust. The Agency finds the results of these studies biologically plausible and their findings reproducible and consistent. It is true that some of these studies, like all human studies, have limitations of sample size, involve confounding exposures, have exposure measurement problems, and often do not produce the kind of dose-response data that can be obtained when experimental animals are subjected to controlled laboratory conditions. What the large group of studies being relied upon by OSHA to establish the significance of the risk associated with exposure to wood dust do show is that the overall weight of evidence that such exposures are harmful and cause loss of functional capacity and material impairment of health is convincing beyond a reasonable doubt.[31]

The EPA has used WOE in the assessment of Superfund sites, endocrine disruptors, and carcinogens. In its 1986 carcinogen assessment guidelines, the EPA introduced the term WOE to describe how it combined tumor findings in animals and humans as the principal elements of its WOE analysis to ascertain the carcinogenicity rating of a compound. In subsequent years, the EPA expanded its framework for a WOE evaluation of carcinogenicity by including a wider range of evidentiary sources beyond rodent and human epidemiological studies. In its recent policy document, "Proposed Guidelines for Carcinogen Risk Assessment"[32] the EPA stated that the agency would include structure-activity relationships (computer models of chemical substances) of other carcinogenic agents, modes of action of carcinogenic agents at cellular and subcellular levels, and knowledge of toxicokinetic and metabolic processes, in addition to the more conventional sources of evidence.

In 1986, the EPA issued a summary ranking of five grades for possible carcinogenic agents (A through E, A signifies that a chemical is a human carcinogen, B a probable human carcinogen, etc., until we get to E, not a carcinogen). In 1996, the EPA replaced the letters with three designations: known/likely a human carcinogen, cannot be determined, and not likely a human carcinogen. The change in the carcinogen guidelines accompanied a more expansive view of the acceptable sources of evidence, which the agency defines as a WOE approach. The EPA referred to a WOE evaluation as a "collective evaluation of all pertinent information so that the full impact of biological plausibility and coherence are adequately considered."[33]

The EPA notes that for a WOE approach, no single "weighing factor" determines the overall weight; moreover, "the factors are not scored numerically by adding pluses and minuses."[34] The factors are judged in combination, and there is no algorithm to aggregate the modalities and quality of evidence. The EPA does provide a guidance document that indicates when the weight goes up or down. Evidence is weighted more highly when time between exposure and outcome is short; there are consistent results in independent studies; a strong association exists between a compound and an effect; there are reliable exposure data; there is a dose-response relationship; there are no biases and confounding factors; there is a high level of statistical significance; and positive results are found in multiple species, sites, and sexes. The agency wrote: "Generally, the weight of human evidence increases with the number of adequate studies that show comparable results on populations exposed to the same agent under different conditions."[35] These qualitative weighting factors are consistent with the Bradford-Hill criteria for inferring causation.[36]

As previously noted, the EPA defined three descriptors for carcinogenicity (I, known/likely; II, cannot be determined; and III, not likely) and asserted that: "Applying a descriptor is a matter of judgment and cannot be reduced to a formula."[37]

What happens when you bring scientists together and ask them to apply a WOE qualitative heuristic and reach a conclusion on whether a substance is, is likely, or is unlikely to be harmful? Several studies have evaluated expert panels' use of WOE to determine whether there is consistency and convergence on the application of the criteria.[38] Some panel studies have introduced weighting factors for specific evidentiary modalities (e.g., in one case, studies that show direct mechanistic evidence for an effect receive a ranking of "1.0," whereas mechanistic data on related compounds receive a ranking of "0.71.") and measured the degree of consensus among experts.[39] The results in the study were mixed. The six teams of experts could not always agree on the direction of the interaction effect of two chemicals after reviewing and ranking the same data and applying the same *a priori* ranking scheme.

One of the key factors behind the reliability of science is the accuracy and replicability of measurement. The term WOE may suggest that a measurement is involved, but that is a false implication of the term. Weighing the evidence, in the way it is carried out by regulatory bodies, is based on human judgment. Such judgments are rarely, if ever, tested for interrater reliability. Those who are considered experts in "weighing" evidence are considered so because they have a good grasp of the type and variety of evidence that, according to standards in their discipline, are sufficient to justify a claim of cause and effect.

WOE IN LEGAL TESTIMONY

In law and public policy, three standards of evidence are generally recognized: preponderance, clear and convincing, and beyond a reasonable doubt. By preponderance of evidence, it is usually meant that a hypothesis under consideration need only be proven more trustworthy (more probable) than its negation. Most civil proceedings use a preponderance of evidence as a standard of proof.

A higher standard is found in the phrase "clear and convincing evidence." The

supporting evidence under this standard has to have more than a marginal edge over the alternative hypothesis. It has been described as evidentiary support "sufficiently strong to command the unhesitating assent of every reasonable mind."[40]

Finally, evidentiary criterion that meets the standard "beyond a reasonable doubt" is the highest burden and the one used in criminal trials to minimize false positives (convicting an innocent person).

In *Daubert v Merrell Dow Pharmaceuticals, Inc,* the US Supreme Court issued a ruling clarifying standards for federal judges on the admissibility of expert testimony in the courtroom. According to the *Daubert* standard, admissible expert testimony must meet a standard of relevancy and reliability. Moreover, some judges apply the standard to each study on which the expert relies, as well as the expert's overall conclusions. This interpretation of *Daubert* would have each study stand on its own. McGarity calls this interpretation of *Daubert* the "corpuscular approach to expert testimony."[41] He writes:

"If the plaintiff fails to establish the relevance and scientific reliability of a sufficient number of individual studies, the trial judge will exclude the expert's testimony and (in the absence of other relevant and reliable expert testimony on causation) grant the defendant's motion for summary judgment before the jury ever enters the picture."[42]

If McGarity is correct on how *Daubert* has been applied, then we will begin to witness a divergence between judicial and regulatory approaches to evidence. In regulation, the strands of evidence are not assumed to stand by themselves. Rather, they are seen as pieces of a puzzle. McGarity notes: "corpuscular approach effectively prevents the expert in toxic tort cases from applying the 'weight-of-evidence' approach that regulatory agencies universally employ in addressing the risks that toxic substances pose to human beings."[43] He likens the WOE approach in risk assessment to the jury's role in civil trials in weighing the quality and credibility of various testimonies.

Because there is no algorithm or canonical methodology for determining WOE in circumstances where no single study is definitive and there is no determinative experiment that can foster a consensus on causality, experts will exercise their judgment on the strength of evidentiary support when a subset of the pieces of the puzzle are assembled. The term puzzle solving is an apt metaphor for the practice of science. Thomas Kuhn used it in his classic study *The Structure of Scientific Revolutions* to describe the role of scientists engaged in normal research problems. "Bringing a normal research problem to a conclusion is achieving the anticipated in a new way, and it requires the solution to all sorts of complex instrumental, conceptual, and mathematical puzzles. The man who succeeds proves himself the expert puzzle-solver."[44] The metaphor has also been cited by Susan Haack in connection with the *Daubert* decision: " . . . scientists are like a bunch of people working, sometimes in cooperation with each other, sometimes in competition, on this or that part of a vast crossword. . . ."[45]

Two experts may easily disagree on the WOE. Who should decide whether the WOE has been met for a given hypothesis when there are contested views? After the corpuscular interpretation of *Daubert*, a judge applies the reliability standard to the admissibility of every piece of evidence in expert testimony without seeing it as part of the entire evidentiary record. By disqualifying the evidence as unreliable on its own weight, jurors may never hear the total weight of scientific evidence. McGarity concludes: "It is not at all clear that lay judges have the wherewithal to distinguish unreliable expert testimony from reliable testimony based on scientific studies that have been 'deconstructed' by paid industry consultants."[46]

When an agency reports, "according to a WOE determination chemical X causes (does not cause) a human disease," a number of possible presuppositions are implicit in the decision process including:

- a socially constructed heuristic for classifying studies or evaluating data,[47]
- an *a priori* numerical weighting scheme, and
- a constructed consensus from a panel of scientists through an interactive consultative process, such as the Delphi Process.

Studies that have measured the variance in expert judgments on the use of WOE in evaluating a hypothesis demonstrate that the application of WOE is not strictly a science but depends on the experience, as well as other tacit factors associated with the expert, such as their familiarity with or financial connection to the substance being evaluated. Experts who apply a WOE analysis to evaluate the human health hazards of a substance draw from their personal knowledge of similar compounds; situate the properties of the compound in a ranking system; and, based on the diversity and quality of the evidence, reach an informed, albeit subjective, judgment on whether the likelihood that the substance is the cause of a human disease is strong, moderate, or weak (e.g., the substance is a human carcinogen, a reproductive toxicant, or an endocrine disruptor).[48] Without an accepted canonical methodology or standard of weighing and combining information streams, and because subjective factors inevitably shape the outcome of the process, judges may not be in any better position than jurors to decide which WOE analysis used by expert witnesses is more credible or reliable.

CONCLUSIONS

As a metaphor, the term WOE turns a cognitive and subjective process, as in the case of juries "weighing the evidence," into something that connotes a purely rational and objective process. If we add the term "scientific" to the phrase, as in "weight of scientific evidence," it suggests even more precision by drawing its symbolic meaning from the terms "weighing" (from the weights and measures) and "science" (the most dependable self-correcting system for fixing belief). In this metaphor there is a triple dose of constructed rationality. Our first realization is that the "weighing instrument" for "weighing evidence" is human cognition, which has never been calibrated to the task. In fact, "weighing evidence" has little if anything in common with weights and measures. Secondly, evidence for a hypothesis generally appears in gradations, with the exception of the evidence from a crucial experiment. Generally, there is more or less evidence or conflicting evidence, or more or less uncertainty in the evidence. The approach that uses WOE applies a

method that treats evidence as a continuous variable and turns it into a dichotomous (below or above the threshold) or triadic variable: "yes," "no," or "probably." (I am indebted to Susan Haack for suggesting this point.) Third, the process of assigning values (qualitative or quantitative) to different evidentiary modalities or to studies of different quality within the same modality is generally constructed *a priori* (independent of empirically based evidence) for each specific case. Where frameworks or models have been developed for this purpose, they have not been standardized.[49]

Writing about the environmental etiology of childhood diseases, Debaun and Gurney highlight the essential role of a conceptual framework for weighing the evidence. "Informed recommendations require systematic assessments of the weight of evidence from available studies and placement of the studies into a conceptual framework that allows for available data to be reviewed in the context of epidemiology principles of causal inference."[50] Presuppositions within these frameworks about the value of different forms of evidence may bias the outcome of a WOE analysis. For example, some WOE approaches give higher weight to mechanistic information over epidemiological data. Where mechanistic knowledge may be unavailable for a particular substance, the value of excellent human epidemiological data may be reduced in the weighing schema because of *a priori* assumptions about evidence.

The use of all the relevant evidence for assessing the health effects of a substance is certainly an advance over restricting assessment to a few choice evidentiary modalities, where information derived from these modalities is scarce or the results highly uncertain. A legal process that rejects the use of WOE or restricts its utilization seems to be at odds with current practices in regulatory science, where knowledge about a potentially hazardous product is pursued through a triangulation of evidentiary streams. Moreover, the same legal processes that acknowledge the value of WOE must also acknowledge that its use is not a rigorous science and, therefore, must be open to public view and interpretation. When WOE is used consistently and uniformly by a regulatory body, it enables that body to develop a strong comparative approach for assessing the potential health and environmental effects of products. On the other hand, the transparency of WOE will enable jurors and stakeholders to fully grasp the norms and *a priori* assumptions that enter into the analysis. The Daubert decision and subsequent related procedures should neither serve as an excuse for "disbarring" WOE analysis in risk assessment nor prevent jurors from learning about the value and limitations that it may bring to litigation. ■

About the Author

The author is with the Department of Urban and Environmental Policy and Planning at Tufts University.

Request for reprints should be sent to Sheldon Krimsky, PhD, Department of Urban and Environmental Policy and Planning, Tufts University, Medford, MA 02155 (e-mail: sheldon.krimsky@tufts.edu).

This article was accepted July 27, 2004.

Acknowledgments

This work was supported in part by the Project on Scientific Knowledge and Public Policy.

Special thanks to the SKAPP Planning Committee, especially David Ozonoff, and participants at the Coronado Conference in 2003 for their constructive comments on an earlier version of the paper.

References

1. A.S. Eddington, *Space, Time and Gravitation.* (Cambridge, UK: Cambridge University Press, 1920).

2. K.R. Popper, *The Logic of Scientific Discovery.* (New York: Harper, 1959), 277.

3. K. Rothman. *Causation and Causal Inference in Epidemiology.* Draft paper delivered to the Coronado Conference on Scientific Evidence and Public Policy, March 3-4, 2003.

4. C.C. Willhite. "Weight-of-Evidence versus Strength-of-Evidence in Toxicologic Hazard Identification: Di(2-Ethylhexyl)Phthalate (DEHP)," *Toxicology* 160 (2001): 219-226.

5. J.M. Culp, R.B. Lowell, and K.J. Cash. "Integrating Mesocosm Experiments with Field and Laboratory Studies to Generate Weight-of-Evidence Risk Assessments for Large Rivers." *Environmental Toxicology and Chemistry* 19 (2000): 1167-1173.

6. L.W. Hall and J.M. Giddings. "The Need for Multiple Lines of Evidence for Predicting Site-Specific Ecological Effects." *Human and Ecological Risk Assessment* 6 (2000): 679-710.

7. C. Menzie, M.H. Henning, J. Cura, et al. "A Weight-of-Evidence Approach for Evaluating Ecological Risks: Report of the Massachusetts Weight-of-Evidence Work Group." *Human Ecological Risk Assessment* 2 (1996): 277-304.

8. G.W. Suter II, ed. *Ecological Risk Assessment.* (Chelsea, MI: Lewis Pub. Co, 1993), 86.

9. V.R. Walker, "Risk Characterization and the Weight of Evidence: Adapting Gatekeeping Concepts from the Courts." *Risk Analysis* 14 (1996): 793-799.

10. G.E. Dallal, Chief, Biostatistics Unit, *The Little Handbook of Statistical Practice* (The Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University), available at http://www.tufts.edu/~gdallal/LHSP.HTM.

11. R.A. Chechile "Probability, utility, and decision trees in environmental decision analysis," in *Environmental Decision Making: A Multidisciplinary Perspective.* (New York: Van Nostrand, 1991), 64-91.

12. Menzie, 1.

13. M.A. Ibrahim, G.G. Bond, T.A. Burke et al. "Weight of the Evidence on the Human Carcinogencitiy of 2,4-D." *Environmental Health Perspectives* 96 (1991): 213-222.

14. R.L. Cooper and R.J. Kavlock. "Endocrine Disruptors and Reproductive Development: A Weight of Evidence Overview. *Journal of Endocrinology* 152 (1997): 159-166.

15. Agency for Toxic Substances and Disease Registry (ATSDR), "The Assessment Process: An Interactive Learning Program," available at http://www.atsdr.cdc.gov/training/public-health-assessment-overview/html/module2/sv18.html. Accessed March 24, 2005.

16. Agency for Toxic Substances and Disease Registry (ATSDR), "The Assessment Process: An Interactive Learning Program," available at http://www.atsdr.cdc.gov/training/public-health-assessment-overview/html/module2/sv18.html. Accessed March 24, 2005.

17. Menzie.

18. A. Edwards, G. Elwyn, K. Hood, and S. Rollnick "Judging the Weight of Evidence in Systematic Reviews: Introducing Rigour into the Qualitative Overview Stage by Assessing Signal and Noise." *Journal of Evaluation in Clinical Practice* 6 (2000): 177-184.

19. R.L. Cooper and R.J. Kavlock "Endocrine Disruptors and Reproductive Development: A Weight-of-Evidence Review." *Journal of Endocrinology* 152 (1997):159-166.

20. C.G. Graves, G.M. Matanoski, and R.G. Tardiff "Weight of Evidence for an Association between Adverse Reproductive and Developmental Effects and Exposure to Disinfection By-Products: A Critical Review." *RegulatoryToxicology and Pharmacology* 34 (2001): 103-124.

21. Ibid, 110.

22. World Health Organization, IPCS Global Assessment of the State of the Science of Endocrine Disruptors: Chapter 7. "Causal Criteria for Assessing Endocrine Disruptors—a Proposed Framework," 123-128 http://www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en/. Accessed March 24, 2005.

23. E.J. Calabrese, L.A. Baldwin, P.T. Kostecki, et al. "A Toxicologically Based Weight-of-Evidence Methodology for the Relative Ranking of Chemicals of Endocrine Disruption Potential." *Regulatory Toxicology and Pharmacology* 26 (1997): 36-40.

24. E.P. Smith, I. Lipkovich, and K. Ye. *Weight of Evidence (WOE): Quantitative Estimate of Probability of Impact.* Working Paper. February 10, 2002.

25. D.J. Balding "When Can a DNA Profile Be Regarded as Unique?" *Science and Justice* 39 (1999): 257-260.

26. I.W. Evett, L.A. Forman, G. Jackson, et al. "DNA Profiling: a Discussion of Issues Relating to the Reporting of Very Small Match Probabilities." *Criminal Law Review* (May 2000): 341-355.

27. International Joint Commission (IJC). *Sixth Biennial Report on Great Lakes Water Quality.* Washington, DC: International Joint Commission, 1992.

28. H.R. Pohl, N. Roney, M. Fay, et al. "Site-Specific Consultation for a Chemical Mixture." *Toxicology and Industrial Health* 15 (1999): 470-479.

29. Agency for Toxic Substances and Disease Registry (ATSDR). *Public Health Assessment Guidance Manual*, Ch. 8, Health effects evaluation: weight-of-evidence analysis, available at http://www.atsdr.cdc.gov/HAC/PHAmanual/ch8p1.html. Accessed June 5, 2004.

30. Agency for Toxic Substances and Disease Registry (ATSDR), Chapter 8, http://www.atsdr.cdc.gov/HAC/PHAmanual/ch8p1.html, pp. 2-3. Accessed March 24, 2005.

31. 29CFR Par. 1910. Air Contaminants, Sec. VI. *Health Effects Discussion and Determination of Final Pel*, January 1989.

32. Environmental Protection Agency. Proposed Guidelines for Carcinogen Risk Assessment. *Federal Register* 61(79):17960-18011 (April 23, 1996). Hereafter, EPA 1996.

33. Ibid., 17981.

34. Ibid.

35. Ibid.

36. A. Bradford-Hill, "The Environment and Disease: Association or Causation?" *Proc. Royal Soc. Med.* 58 (1965): 295-300.

37. EPA 1996, 17985.

38. M.E. Anderson, M.E. Meek, G.A. Boorman, et al. "Lessons Learned in Applying the U.S. EPA Proposed Cancer Guidelines to Specific Compounds." *Toxicological Sciences* 53 (2000): 159-172.

39. M.M. Muntaz, P. Furkin, G.I. Diamond, et al. "Exercises in the Use of Weight-of-Evidence Approach for Chemical-Mixture Interactions." *Journal of Clean Technology, Environmental Toxicology, and Occupational Medicine* 5 (1996): 339-345.

40. V.R. Walker "Risk Characterization and the Weight of Evidence: Adapting Gatekeeping Concepts from the Courts." *Risk Analysis* 16 (1996): 793-799.

41. T.O. McGarity "Proposal for Linking Culpability and Causation to Ensure Corporate Accountability for Toxic Risks." *William and Mary Environmental Law and Policy Review*, Fall 2001.

42. McGarity, 7.

43. Ibid., 8.

44. T.S. Kuhn. *The Structure of Scientific Revolutions.* (Chicago: University of Chicago Press, 1962), 36.

45. S. Haack "An Epistemologist in the Bramble-Bush: at the Supreme Court with Mr. Joiner." *Journal of Health, Politics, Policy & Law,* 26 (2001): 217-248.

46. McGarity, 12.

47. A Baysian statistical approach to WOE is vien in: E.P. Smith, I. Lipkovich, and K. Ye. *Weight of evidence (WOE): Quantitative estimation of probability impact.* Unpublished manuscript. (Blacksburg, VA: Department of Statistics, Virginia Tech., 2002).

48. The ATSDR endorses "the use of a narrative statement incorporating "weight-of-evidence" conclusions in lieu of alphanumeric designations alone in conveying qualitative conclusions regarding carcinogenicity," available at: http://www/atsdr.cdc.gov/cancer.html, p. 9. Accessed March 24, 2005.

49. Op. cit. Ibrahim et al. 1991, p. 219.

50. M.R. DeBaun and J.G. Gurney "Environmental Exposure and Cancer in Children: a Conceptual Framework for the Pediatrician." *Pediatric Clinics of North America* 48 (2001): 1125.

Exhibit E

NIH Public Access
Author Manuscript
Obstet Gynecol. Author manuscript; available in PMC 2013 September 24.

Published in final edited form as:
Obstet Gynecol. 2011 May ; 117(5): 1042–1050. doi:10.1097/AOG.0b013e318212fcb7.

Assessing Ovarian Cancer Risk When Considering Elective Oophorectomy at the Time of Hysterectomy

Allison F. Vitonis, SM, **Linda Titus-Ernstoff, PhD**, and **Daniel W. Cramer, MD, ScD**
Department of Obstetrics and Gynecology, Epidemiology Center, Brigham and Women's Hospital, Boston, Massachusetts; and Norris Cotton Cancer Center, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire

Abstract

Objective—To develop a risk-factor score that may provide additional guidance to women and their physicians regarding elective bilateral salpingo-oophorectomy at the time of hysterectomy.

Methods—From a case–control study conducted from 1992 to 2008 in women residing in eastern Massachusetts or New Hampshire, we selected 1,098 women with invasive ovarian cancer (case group) and 1,363 for the control group who were older than 40 years and had neither hysterectomy nor a personal or family history of breast or ovarian cancer. Using logistic regression, we identified key risk factors and built a risk score. The score was separately assessed in 126 women in the case group and 156 in the control group with excluded prior hysterectomy to determine whether women who developed ovarian cancer could have been distinguished.

Results—Summing eight conditions found to be associated with ovarian cancer (Jewish ethnicity, less than 1 year of oral contraceptive use, nulliparity, no breastfeeding, no tubal ligation, painful periods or endometriosis, polycystic ovary syndrome or obesity, talc use), we created a five-level score. Assigning average risk to those with a score of 2, the odds ratios varied from 0.56 (95% confidence interval [CI] 0.42–0.74) for a score of 0–1 to 3.30 (95% CI 2.50–4.35) for a score of 5 or greater (*P* trend <.001). The risk score was higher for women who developed ovarian cancer after hysterectomy than those who did not (*P*=.01). Lifetime risks for ovarian cancer for a woman at age 40 years are changed from 1.2% with a 0–1 score to 6.6% with a score of 5 or higher.

Conclusion—We developed a risk-assessment tool that can quantify women's risk for ovarian cancer and should be validated in other data sets.

In the 1990s, there were approximately 600,000 hysterectomies performed in the United States annually and 55% of these also involved bilateral salpin-go-oophorectomy,[1] many done solely to reduce the risk for ovarian cancer. It has been suggested that elective bilateral salpingo-oophorectomy be considered for women older than 40 years,[2–4] whereas surveys in the United Kingdom revealed that 85–90% of physicians recommended bilateral salpingo-oophorectomy for postmenopausal women coming to hysterectomy.[5,6] However, Parker et al,[7] citing evidence that postmenopausal ovaries secrete androgens important to health, performed a risk–benefit analysis and concluded that ovarian conservation benefits long-term survival for women at "average risk" for ovarian cancer undergoing hysterectomy for benign disease. A subsequent study using observational data from the Nurses' Health Study

© 2011 by The American College of Obstetricians and Gynecologists

Corresponding author: Daniel W. Cramer, MD, Brigham and Women's Hospital, 221 Longwood Avenue, RFB 365, Boston, MA 02115; dcramer@partners.org.

Financial Disclosure: The authors did not report any potential conflicts of interest.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

on all and various causes of mortality for hysterectomized women with and without oophorectomy supported their conclusion.[8]

In addressing the value of bilateral salpingo-oophorectomy, Parker et al distinguished average-risk women from those with known *BRCA1* or *BRCA2* mutations or a strong family history of breast and ovarian cancer. In the latter group, bilateral salpingo-oophorectomy may truly be beneficial in reducing risk for both breast and ovarian cancer.[9] Genetic or familial risk factors or both, however, account for a small proportion of ovarian cancer. Consequently, it is important to assess ovarian cancer risk among women who lack the genetic or familial profile. In this article, we describe a risk-factor score that may be of value in further categorizing risk for ovarian cancer in women without a personal or family history of cancer to provide additional guidance to women and their physicians regarding elective bilateral salpingo-oophorectomy at the time of hysterectomy.

Materials and Methods

Data used in this study come from three enrollment phases of a case–control study of ovarian cancer in New England. The earlier two phases have been described previously.[10] Briefly, we used statewide cancer registries and hospital tumor boards to identify ovarian cancer cases diagnosed in eastern Massachusetts and the entire state of New Hampshire from May 1992 to March 1997 and August 1998 to April 2003. We enrolled 1,306 women in the case group of whom 1,231 had been diagnosed with epithelial ovarian cancers. Women for the control group for the first phase of the study were identified by random-digit dialing supplemented with residents' lists for older control-group participants. Approximately 10% of households randomly dialed contained an eligible control and of these, 421 (72%) agreed to participate. All women for the control group for the second phase were identified through town resident lists (town books) in Massachusetts and drivers' license registries in New Hampshire. Of the 2,102 potential control-group participants identified through town books in both phases, 635 were ineligible, 644 declined participation, and 823 were enrolled. In total, 1,244 women were enrolled in the control group.

In the third enrollment phase, between October 2003 and November 2008, we identified 1,610 women residing in eastern Massachusetts or New Hampshire who had a diagnosis of incident ovarian cancer. Of these, 372 could not be contacted because they had died, moved, or had no telephone; did not speak English; had a nonovarian primary tumor after review; or lived outside the study area. Physicians declined permission to contact 128, and 213 declined or were too ill to participate. The remaining 897 women were enrolled; of these, 845 had epithelial ovarian tumors, including tumors of borderline malignancy.

Similar to the second phase of the study, control-group participants were identified through town books in Massachusetts and drivers' license lists in New Hampshire. Age matching was accomplished by sampling control-group participants based on the age distribution of women in the case group in the previous phases of the study with adjustment as current case-group participants were enrolled. Of the 2,523 potential control-group participants identified, 850 were ineligible because they had died, moved, had no telephone, did not speak English, had no ovaries, or were seriously ill and 816 potential control-group participants declined participation either by phone or by “opt out” postcard. A total of 857 control-group participants were enrolled.

In all phases, after written informed consent, demographic information, reproductive and medical history, and habits were assessed by in-person interview. All of the questions were framed with respect to a reference date defined as 1 year before the diagnosis date for

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

women in the case group and the date of interview for those in the control group. Histologic type, grade, and stage of disease were abstracted from case pathology reports. This study was approved by Brigham and Women's Hospital and Dartmouth Medical Center's institutional review boards.

We used two approaches to identify women who may be at greater risk for ovarian cancer after hysterectomy and more likely to benefit from elective bilateral salpingo-oophorectomy. In the first approach, we constructed a risk-factor score that would be relevant to decision-making for “average-risk” women coming to hysterectomy. For this analysis, we excluded all women who had prior hysterectomy (n=368). We also excluded women who would be deemed to be at above-average risk because of a personal history of breast cancer or a family history of ovarian cancer at any age or breast cancer diagnosed before age 50 years (n=532). We excluded women younger than 40 years because they would unlikely be offered bilateral salpingo-oophorectomy without an indication (n=615). We also restricted the analysis to women who had an invasive ovarian cancer, whose survival is substantially worse compared with those with borderline tumors. The final sample included 1,098 women in the case group (including 17 primary peritoneal cases) and 1,363 in the control group. In the second approach, also after excluding borderline cases and women with a personal or family history of breast or ovarian cancer, we examined women in the case (n=126, including one primary peritoneal case) and control groups (n=156) who had previous hysterectomy to determine whether risk profiles or reasons for the surgery could have distinguished women who subsequently developed ovarian cancer.

In both approaches, unconditional logistic regression models were used to identify significant risk factors distinguishing women in the case group from those in the control group. Continuous variables were categorized based on quartiles of the control distributions. Associations are presented as odds ratios, 95% confidence intervals, and Wald test *P* values. We used Wald tests from logistic regression to test for trends in ordinal categorical exposures by creating ordinal variables in which the median value or midpoint of each category was assigned to all participants within that category. To evaluate whether associations between risk factors and ovarian cancer varied by study phase, we conducted stratified analyses and likelihood ratio tests comparing models with both main effects and interaction terms with models with main effects only. Because of the small amount of missing data in this study, participants with missing exposures were dropped from analyses. Combinations of factors were examined to identify the best cumulative index of experiences associated with ovarian cancer risk. In all models, we adjusted for study phase and the matching variables age (continuous) and study site (Massachusetts, New Hampshire).

We translated the relative risks obtained from the model into absolute risks by multiplying them by cumulative risks for ovarian cancer occurrence with age 85 years as an end point. Cumulative risks were calculated from 2003–2007 age-specific incidence rates for ovarian cancer provided through Surveillance, Epidemiology, and End Results (SEER) of the National Cancer Institute.[11] These rates are based on all women as the denominator including women with an oophorectomy, whereas we wish cumulative risk to apply only to women with intact ovaries. From a study that examined the effect of hysterectomy and oophorectomy on genital cancer rates,[12] we adjusted age-specific incidence rates upward based on estimates of the prevalence of oophorectomy by dividing each age-specific incidence rate by one minus the prevalence of oophorectomy in that age group. Cumulative incidence was calculated by summing the adjusted age-specific incidence rates times the duration of the age-specific incidence intervals as described in Rothman and Greenland.[13]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Results

Table 1 shows the distribution of women in the case and control groups by study details and well-established or potential risk factors for ovarian cancer. The majority of women enrolled in the case group were white, which limited our ability to include race as a risk factor. We observed highly significant increases in risk associated with lack of oral contraceptive use, nulliparity, never having breastfed, no tubal ligation, painful periods or endometriosis, polycystic ovarian syndrome or obesity (body mass index [calculated as weight (kg)/[height $(m)]^2$] greater than 30), and long-term genital talc use. An increasing number of estimated ovulatory cycles not interrupted by pregnancies, breastfeeding, or oral contraceptive use was also strongly associated with increased risk. Having a Jewish ethnic background was associated with increased risk but of borderline significance (P=.08). There was no significant association with age at menarche or menopause, fertility hormones, or menopausal hormone use (except for progesterone-only regimens, which were used by few participants in this study). We observed no significant interactions between risk factors and study phase.

The final entry in Table 1 shows the results of a simple score created to summarize risk by number of ovarian cancer risk factors. Conditions included in this score are Jewish ethnicity, more than 1 year of oral contraceptive use, nulliparity, no breastfeeding, no tubal ligation, painful periods or endometriosis, polycystic ovarian syndrome or obesity, and long-term genital talc use. There was a significant trend of increasing risk with increasing number of conditions (P trend <.001). Compared with women with two conditions, women with zero to one condition had a 40% reduction in risk, whereas women with three, four, and five or more conditions had 60%, twofold, and threefold increases in risk, respectively. We examined this score by histologic subtype and stage of invasive epithelial ovarian cancer and observed significant trends of increasing risk for all subtypes and early- and late-stage disease (Table 2).

Table 3 shows the results of the analysis of ovarian cancer in women in the case and control groups who had prior hysterectomy. There were significant trends for risk of ovarian cancer to be lower with an older age at hysterectomy and greater with a longer interval since performance of the hysterectomy. The most common reasons for hysterectomy (by the woman's self-report) were heavy bleeding, leiomyomas, or both, which were diagnosed in 61.9% of women in the case groups and 57.0% of those in the control group. Compared with this group, there was a lower likelihood for developing ovarian cancer if the reported diagnosis was prolapse (P=.06). Risk of ovarian cancer among hysterectomized women increased monotonically with a higher risk-factor score (P trend=.01). The average risk-factor score was 3.4 for all women in the case group compared with 3.0 for all women in the control group (P=.009) and 3.4 for women in the case group compared with 2.6 for those in the control group (P=.01) for women who underwent hysterectomy after age 45 years. Women with ovarian cancer who had prior hysterectomy had a higher frequency of serous histologic types (67%) and lower frequency of endometrioid and clear cell types (22%) compared with nonhysterectomized women in the case group, in which the respective frequencies were 52% and 36% (P<.001) (data not shown).

Table 4 translates the risk-factor score from Table 1 into absolute risks for the occurrence of ovarian cancer during the remaining years of life from a particular starting age beginning at age 40 years until age 85 years as an end point. Assuming that the category of two risk factors best represents risk in the general population (and therefore the referent category), we multiplied the cumulative risks by 0.6, 1.6, 2.1, and 3.3 for the score categories 0–1, 3, 4, and 5 or more, respectively. As illustrated in Table 4, a woman who is 40 years old and has zero to one risk factors would have an absolute risk of developing ovarian cancer by age 85

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

years of 1.2% (95% CI 0.8–1.4%), whereas a woman with five or more risk factors would have a risk of 6.6% (95% CI 5.0–8.6%).

Discussion

Current American College of Obstetricians and Gynecologist guidelines[14] recommend that family history, menopausal status, and pelvic disease that might predispose to reoperation be considered in whether bilateral salpingo-oophorectomy should be offered to women coming to hysterectomy. The guidelines state that "Strong consideration should be made for retaining normal ovaries in premenopausal woman who are not at increased genetic risk of ovarian cancer." Bilateral salpingo-oophorectomy should be offered to women with known or suspected *BRCA1* or *BRCA2* mutations after completion of childbearing. For postmenopausal women (with normal ovaries), the guidelines state: "Given the risk of ovarian cancer in postmenopausal women, ovarian removal at the time of hysterectomy should be considered for these women." Nulligravidity and family history of ovarian cancer are mentioned as increasing risk for ovarian cancer; and pregnancy, tubal ligation, and use of oral contraceptive are mentioned as decreasing risk. However, no concrete rules are offered on how these characteristics might be used to weigh risk in an individual woman.

In this article, we derive a simple score to help physicians and women weigh individual risk for ovarian cancer. We first excluded those women who would already be viewed at high risk such as those with a personal history of breast cancer or family history (of a mother or sister) with breast cancer (before age 50 years) or ovarian cancer at any age. To make the model most relevant to women considering oophorectomy at the time of hysterectomy, we then excluded women younger than 40 years, who may be inappropriate candidates for elective oophorectomy without known ovarian pathology, as well as women in the case and control groups who had prior hysterectomy. We identified those risk factors to be considered: parity, oral contraceptive use, breastfeeding, tubal ligation, painful periods or endometriosis, obesity or polycystic ovarian syndrome, and talc use. These risk factors are concordant with published epidemiologic data related to reproductive factors,[15–23] use of talc,[17–19] tubal ligation,[20,24–27] endometriosis,[28] and polycystic ovarian syndrome or obesity.[29,30] It is also known that approximately 2% of Jewish women carry one of three founder mutations of *BRCA1* or *BRCA2*. Approximately 40% of Jewish women who present with ovarian cancer will carry a founder mutation.[31] Even after removing those with a family history of breast or ovarian cancer, women with Jewish ethnic backgrounds remain at approximately a 30% increased in risk for ovarian cancer.

Creating simple dichotomies from these factors and summing them allowed a five-level risk score to be constructed, which correlated directly with increasing relative risks for ovarian cancer. Combining various risk factors to create a risk score for ovarian cancer has been performed in studies that have looked at the estimated number of ovulatory cycles, which also directly correlates with ovarian cancer risk.[10,32] However, we did not include ovulatory cycles in our model because estimating them would require a calculator or paper and pencil. Thus, a simple linear combination of diverse risk factors, even those that do not logically fit into an ovulatory cycles score, adds cumulatively to increase ovarian cancer risk. We previously have discussed the potential basis for this phenomenon as indicating a common pathway for many ovarian cancer risk factors operating through their ability to affect immunity related to important cell surface glycoproteins, know as mucins, especially MUC1.[33]

We also performed an analysis on women who had previous hysterectomy. Most case–control studies of ovarian cancer allow women with hysterectomy to be included in the control group as long as they said their operation did not include oophorectomy. Nearly all

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

hysterectomized women who later developed ovarian cancer would be correct in their recollection that they did not have a bilateral salpingo-oophorectomy. However, there is a greater likelihood that those who did not develop ovarian cancer may have incorrectly stated their ovaries were left, leading to misclassification. We are uncertain whether this may partially explain the greater percentage of control-group participants who reported hysterectomy without bilateral salpingo-oophorectomy after age 46 years compared with women in the case group observed in this study. Because historical medical records could not be retrieved for participants, it was also necessary to rely on the woman's recollection of why the surgery was performed. Women who went on to develop ovarian cancer after hysterectomy were less likely to have had hysterectomy for prolapse (P=.06). Regarding our risk score, we again found a significant trend for a higher cumulative score to predict greater risk for ovarian cancer occurring after hysterectomy. Notably, the average score for women who had hysterectomy after age 45 years and subsequently developed ovarian cancer was 3.4 for women in the case group compared with 2.6 for those in the control group (P=.01). It may be particularly important to initiate a dialogue about ovarian cancer risk factors before hysterectomy after this age.

Potential weaknesses of this study derive from the fact that case–control data were used to create our scoring system. Biases may occur in case–control studies that can affect risk estimates, including recall bias leading to misclassification of exposure. In addition, selection biases may occur in that exposures for women with rapidly fatal disease who could not be interviewed may be underrepresented. Nevertheless, the risk factors we observed agree with published data, some of which come from cohort studies in which these biases are less likely to occur and our scoring system was applied to both early- and late-stage disease (Table 2). Another limitation of case–control data is that it allows only relative, not absolute, risks to be calculated directly. To overcome this limitation, we multiplied the odds ratio for each score by estimated lifetime risks of ovarian cancer. The age-specific incidence rates used to calculate lifetime risks were first adjusted upward based on the prevalence of oophorectomy in the general population.

Our risk score does not provide a precise formula for when elective oophorectomy should be recommended because we did not perform a cost–benefit analysis taking into consideration the competing risks from long-term complications of bilateral salpingooophorectomy, including bone fracture and cardiovascular diseases. Based on the rarity of ovarian cancer relative to other conditions considered by Parker et al in their analysis of the Nurses' Health Data, it is possible that, even if all cases of ovarian cancer could be predicted and eliminated, overall benefits might not be shifted toward selective bilateral salpingo-oophorectomy. Nevertheless, we think it is important for physicians and their patients to weigh individual risk for ovarian cancer when considering elective oophorectomy and have a discussion about individual risk for ovarian cancer. Even if the woman at elevated risk elects to conserve ovaries, bilateral salpingectomy without oophorectomy might be considered. Emerging evidence suggests that many high-grade invasive ovarian cancers may have their origin in the fallopian tubes rather than ovaries,[34] prompting Canadian health officials in British Columbia to urge gynecologists to perform salpingectomy (without oophorectomy) on women coming for hysterectomy. Our risk score might enable selection of women who would be candidates for this surgical alternative to oophorectomy if women at higher risk do not elect to have oophorectomy. Although we believe our scoring system is an improvement over existing methods for assessing risk for ovarian cancer in women without a family history, it should be viewed as a prototype until it can be validated in other data sets, especially with prospectively collected data from women including more nonwhites who were underrepresented in our study.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Acknowledgments

Supported by the National Cancer Institute grants to D.W.C.: Ovarian Cancer SPORE P50 CA105009 and R01 CA54419.

References

1. Keshavarz H, Hillis S, Kieke B, Marchbanks PA. Hysterectomy surveillance–United States, 1994–1999. MMWR CDC Surveill Summ. 2002; 51:1–8.
2. Coukos G, Rubin SC. Prophylactic oophorectomy. Best Pract Res Clin Obstet Gynaecol. 2002; 16:597–609. [PubMed: 12413937]
3. Kontoravdis A, Kalogirou D, Antoniou G, Kontoravdis N, Karakitsos P, Zourlas PA. Prophylactic oophorectomy in ovarian cancer prevention. Int J Gynaecol Obstet. 1996; 54:257–62. [PubMed: 8889634]
4. Studd J. Does retention of the ovaries improve long-term survival after hysterectomy? Prophylactic oophorectomy. Climacteric. 2006; 9:164–6. [PubMed: 16766429]
5. Geary M, Geoghegan A, Foley M. Prevention of ovarian cancer: a survey of the practice of prophylactic oophorectomy by consultant gynaecologists in Ireland. Ir Med J. 1997; 90:186–7. [PubMed: 9345830]
6. Jacobs I, Oram D. Prevention of ovarian cancer: a survey of the practice of prophylactic oophorectomy by fellows and members of the Royal College of Obstetricians and Gynaecologists. Br J Obstet Gynaecol. 1989; 96:510–5. [PubMed: 2757978]
7. Parker WH, Broder MS, Liu Z, Shoupe D, Farquhar C, Berek JS. Ovarian conservation at the time of hysterectomy for benign disease. Clin Obstet Gynecol. 2007; 50:354–61. [PubMed: 17513923]
8. Parker WH, Broder MS, Chang E, Feskanich D, Farquhar C, Liu Z, et al. Ovarian conservation at the time of hysterectomy and long-term health outcomes in the nurses' health study. Obstet Gynecol. 2009; 113:1027–37. [PubMed: 19384117]
9. Kauff ND, Barakat RR. Risk-reducing salpingo-oophorectomy in patients with germline mutations in BRCA1 or BRCA2. J Clin Oncol. 2007; 25:2921–7. [PubMed: 17617523]
10. Terry KL, De Vivo I, Titus-Ernstoff L, Shih MC, Cramer DW. Androgen receptor cytosine, adenine, guanine repeats, and haplotypes in relation to ovarian cancer risk. Cancer Res. 2005; 65:5974–81. [PubMed: 15994977]
11. Altekruse, SF.; Kosary, CL.; Krapcho, M.; Neyman, N.; Aminou, R.; Waldron, W., et al. SEER cancer statistics review, 1975–2007. Bethesda (MD): National Cancer Institute; 2010. Available at: http://seer.cancer.gov/csr/1975_2007/ based on November 2009 SEER data submission. Retrieved January 18, 2011
12. Merrill RM. Impact of hysterectomy and bilateral oophorectomy on race-specific rates of corpus, cervical, and ovarian cancers in the United States. Ann Epidemiol. 2006; 16:880–7. [PubMed: 17027290]
13. Rothman, KJ.; Greenland, S. Modern epidemiology. 2nd. Philadelphia (PA): Lippincott Williams & Wilkins; 1998.
14. Obstet Gynecol. Vol. 111. ACOG Practice Bulletin No. 89. American College of Obstetricians and Gynecologists; 2008. Elective and risk-reducing salpingo-oophorectomy; p. 231-41.
15. Whittemore AS, Harris R, Itnyre J. Characteristics relating to ovarian cancer risk: collaborative analysis of 12 US case-control studies. II. Invasive epithelial ovarian cancers in white women. Collaborative Ovarian Cancer Group. Am J Epidemiol. 1992; 136:1184–203. [PubMed: 1476141]
16. Kumle M, Weiderpass E, Braaten T, Adami HO, Lund E. Risk for invasive and borderline epithelial ovarian neoplasias following use of hormonal contraceptives: the Norwegian-Swedish Women's Lifestyle and Health Cohort Study. Br J Cancer. 2004; 90:1386–91. [PubMed: 15054460]
17. Cramer DW, Liberman RF, Titus-Ernstoff L, Welch WR, Greenberg ER, Baron JA, et al. Genital talc exposure and risk of ovarian cancer. Int J Cancer. 1999; 81:351–6. [PubMed: 10209948]
18. Cramer DW, Welch WR, Scully RE, Wojciechowski CA. Ovarian cancer and talc: a case-control study. Cancer. 1982; 50:372–6. [PubMed: 7083145]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

19. Gates MA, Tworoger SS, Terry KL, Titus-Ernstoff L, Rosner B, De Vivo I, et al. Talc use, variants of the GSTM1, GSTT1, and NAT2 genes, and risk of epithelial ovarian cancer. Cancer Epidemiol Biomarkers Prev. 2008; 17:2436–44. [PubMed: 18768514]
20. Tworoger SS, Fairfield KM, Colditz GA, Rosner BA, Hankinson SE. Association of oral contraceptive use, other contraceptive methods, and infertility with ovarian cancer risk. Am J Epidemiol. 2007; 166:894–901. [PubMed: 17656616]
21. Titus-Ernstoff L, Rees JR, Terry KL, Cramer DW. Breastfeeding the last born child and risk of ovarian cancer. Cancer Causes Control. 2010; 21:201–7. [PubMed: 19902367]
22. Purdie DM, Siskind V, Bain CJ, Webb PM, Green AC. Reproduction-related risk factors for mucinous and nonmucinous epithelial ovarian cancer. Am J Epidemiol. 2001; 153:860–4. [PubMed: 11323316]
23. Danforth KN, Tworoger SS, Hecht JL, Rosner BA, Colditz GA, Hankinson SE. Breastfeeding and risk of ovarian cancer in two prospective cohorts. Cancer Causes Control. 2007; 18:517–23. [PubMed: 17450440]
24. Green A, Purdie D, Bain C, Siskind V, Russell P, Quinn M, et al. Tubal sterilisation, hysterectomy and decreased risk of ovarian cancer. Survey of Women's Health Study Group. Int J Cancer. 1997; 71:948–51. [PubMed: 9185694]
25. Hankinson SE, Hunter DJ, Colditz GA, Willett WC, Stampfer MJ, Rosner B, et al. Tubal ligation, hysterectomy, and risk of ovarian cancer. A prospective study. JAMA. 1993; 270:2813–8. [PubMed: 8133619]
26. Ness RB, Grisso JA, Cottreau C, Klapper J, Vergona R, Wheeler JE, et al. Factors related to inflammation of the ovarian epithelium and risk of ovarian cancer. Epidemiology. 2000; 11:111–7. [PubMed: 11021606]
27. Purdie D, Green A, Bain C, Siskind V, Ward B, Hacker N, et al. Reproductive and other factors and risk of epithelial ovarian cancer: an Australian case-control study. Survey of Women's Health Study Group. Int J Cancer. 1995; 62:678–84. [PubMed: 7558414]
28. Vlahos NF, Kalampokas T, Fotiou S. Endometriosis and ovarian cancer: a review. Gynecol Endocrinol. 2010; 26:213–9. [PubMed: 19718562]
29. Chittenden BG, Fullerton G, Maheshwari A, Bhattacharya S. Polycystic ovary syndrome and the risk of gynaecological cancer: a systematic review. Reprod Biomed Online. 2009; 19:398–405. [PubMed: 19778486]
30. Olsen CM, Green AC, Whiteman DC, Sadeghi S, Kolahdooz F, Webb PM. Obesity and the risk of epithelial ovarian cancer: a systematic review and meta-analysis. Eur J Cancer. 2007; 43:690–709. [PubMed: 17223544]
31. Robles-Diaz L, Goldfrank DJ, Kauff ND, Robson M, Offit K. Hereditary ovarian cancer in Ashkenazi Jews. Fam Cancer. 2004; 3:259–64. [PubMed: 15516850]
32. Tung KH, Wilkens LR, Wu AH, McDuffie K, Nomura AM, Kolonel LN, et al. Effect of anovulation factors on pre- and postmenopausal ovarian cancer risk: revisiting the incessant ovulation hypothesis. Am J Epidemiol. 2005; 161:321–9. [PubMed: 15692075]
33. Cramer DW, Titus-Ernstoff L, McKolanis JR, Welch WR, Vitonis AF, Berkowitz RS, et al. Conditions associated with antibodies against the tumor-associated antigen MUC1 and their relationship to risk for ovarian cancer. Cancer Epidemiol Biomarkers Prev. 2005; 14:1125–31. [PubMed: 15894662]
34. Salvador S, Gilks B, Kobel M, Huntsman D, Rosen B, Miller D. The fallopian tube: primary site of most pelvic high-grade serous carcinomas. Int J Gynecol Cancer. 2009; 19:58–64. [PubMed: 19258943]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Conditions and Exposures Associated With Invasive Ovarian Cancer

	No. of Women in the Case Group (n = 1,098)	No. of Women in the Control Group (n = 1,363)	OR (95% CI)*	*P*
Study				
Phase 1: 1992–1997	284 (25.9)	316 (23.2)		
Phase 2: 1998–2003	327 (29.8)	456 (33.5)		
Phase 3: 2003–2008	487 (44.4)	591 (43.4)		
Site				
Massachusetts	860 (78.3)	1,117 (82.0)		
New Hampshire	238 (21.7)	246 (18.0)		
Race				
White	1,056 (96.2)	1,335 (98.0)	1.00	
African American	15 (1.4)	11 (0.8)	1.79 (0.82–3.93)	.15
Hispanic	9 (0.8)	12 (0.9)	1.02 (0.42–2.43)	.97
Asian	14 (1.3)	3 (0.2)	6.28 (1.80–21.9)	.004†
Other	4 (0.4)	2 (0.2)	2.52 (0.46–13.8)	.29
Jewish ethnicity				
No	1,017 (92.6)	1,283 (94.1)	1.00	
Yes	81 (7.4)	80 (5.9)	1.34 (0.97–1.85)	.08
Oral contraceptive use				
1 y or more	436 (39.7)	726 (53.3)	1.00	
Less than 1 y or no use	662 (60.3)	637 (46.7)	1.81 (1.52–2.15)	<.001
Parity				
Parous	794 (72.3)	1,162 (85.2)	1.00	
Nulliparous	304 (27.7)	201 (14.8)	2.34 (1.91–2.87)	<.001
Breastfeeding				
Any	353 (32.2)	690 (50.6)	1.00	
None	745 (67.8)	673 (49.4)	2.18 (1.84–2.57)	<.001
Tubal ligation				
Yes	142 (12.9)	294 (21.6)	1.00	
No	956 (87.1)	1,069 (78.4)	1.87 (1.50–2.33)	<.001
Pain with periods or endometriosis				
No	642 (58.5)	925 (67.9)	1.00	
Yes	456 (41.5)	438 (32.1)	1.53 (1.30–1.81)	<.001
PCOS or obesity (BMI more than 30 kg/m^2)				
No	785 (71.5)	1,039 (76.2)	1.00	
Yes	313 (28.5)	324 (23.8)	1.27 (1.06–1.52)	.01
Long-term genital talc use (10 y or more)				
No	932 (84.9)	1,211 (88.8)	1.00	
Yes	166 (15.1)	152 (11.2)	1.42 (1.12–1.81)	.004
Ovulatory cycles				
Quartile 1	149 (14.5)	317 (25.0)	1.00	

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

	No. of Women in the Case Group (n = 1,098)	No. of Women in the Control Group (n = 1,363)	OR (95% CI)*	*P*
Quartile 2	218 (21.2)	316 (24.9)	1.51 (1.16–1.97)	.002
Quartile 3	300 (29.1)	317 (25.0)	2.14 (1.65–2.77)	<.001
Quartile 4	363 (35.2)	319 (25.1)	2.63 (2.02–3.43)	<.001
Early menarche (younger than 12 y)				
Younger than 12	237 (21.7)	283 (20.8)	1.03 (0.85–1.25)	.77
12–15	815 (74.5)	1,006 (74.0)	1.00	
Older than 15	42 (3.8)	71 (5.2)	0.73 (0.49–1.08)	.11
Age at natural menopause (y)				
Younger than 49	243 (33.2)	283 (33.0)	1.00	
49–51	228 (31.1)	272 (31.7)	0.99 (0.77–1.27)	.93
Older than 51	262 (35.7)	303 (35.3)	1.03 (0.81–1.32)	.80
Postmenopausal hormone use				
None	839 (76.8)	983 (72.6)	1.00	
Estrogen only	54 (5.0)	77 (5.7)	0.77 (0.54–1.12)	.18
Estrogen and progesterone	174 (15.9)	245 (18.1)	0.83 (0.66–1.03)	.10
Progesterone only	4 (0.4)	17 (1.2)	0.28 (0.09–0.84)	.02
Oral contraceptives	3 (0.3)	8 (0.6)	0.46 (0.12–1.73)	.25
Other	18 (1.6)	25 (1.8)	0.82 (0.44–1.52)	.52
Fertility hormones				
No	1,014 (92.4)	1,255 (92.1)	1.00	
Yes	84 (7.6)	108 (7.9)	0.97 (0.72–1.31)	.84
Total number of risk factors‡				
0–1	98 (8.9)	311 (22.8)	0.56 (0.42–0.74)	<.001
2	201 (18.3)	361 (26.5)	1.00	
3	312 (28.4)	340 (24.9)	1.66 (1.31–2.09)	<.001
4	255 (23.2)	222 (16.3)	2.10 (1.64–2.70)	<.001
5 or more	232 (21.1)	129 (9.5)	3.30 (2.50–4.35)	<.001

OR, odds ratio; CI, confidence interval; PCOS, polycystic ovarian syndrome; BMI, body mass index.

Data are n (%) unless otherwise specified.

* Adjusted for study center, reference age, and study phase.

† The excess of Asian ovarian cancer cases simply may reflect limited ability to recruit Asian women for the control group.

‡ Risk factors include Jewish ethnicity, less than 1 year of oral contraceptive use, nulliparity, no breastfeeding, no tubal ligation, painful periods or endometriosis, PCOS or BMI greater than 30 kg/m^2, and long-term talc use.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 2

Cumulative Index of Experiences Associated With Invasive Ovarian Cancer by Histologic Type and Stage

Total No. of Risk Factors*	Serous Invasive (n=566)†	Mucinous (n=62)†	Endometrioid (n=223)†	Clear Cell (n=175)†	Other or Undifferentiated (n=72)†	Early Stage (I–II) (n=462)†	Late Stage (III–IV) (n=634)†
0–1	0.56 (0.39–0.80)	0.61 (0.24–1.56)	0.66 (0.36–1.20)	0.34 (0.16–0.71)	0.88 (0.32–2.40)	0.52 (0.33–0.82)	0.58 (0.42–0.81)
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	1.39 (1.04–1.84)	1.62 (0.79–3.33)	2.38 (1.50–3.77)	1.43 (0.88–2.33)	3.56 (1.66–7.63)	2.11 (1.51–2.96)	1.43 (1.09–1.87)
4	1.65 (1.22–2.24)	1.46 (0.65–3.27)	3.02 (1.86–4.89)	2.92 (1.82–4.68)	2.95 (1.28–6.83)	3.28 (2.32–4.63)	1.55 (1.16–2.09)
5 or more	2.75 (1.98–3.82)	2.01 (0.86–4.75)	5.78 (3.54–9.42)	3.54 (2.11–5.93)	3.16 (1.25–7.99)	5.17 (3.58–7.47)	2.39 (1.73–3.30)
P trend	<.001	.01	<.001	<.001	<.001	<.001	<.001

Data are odds ratio (95% confidence interval) unless otherwise specified.

*Risk factors include Jewish ethnicity, less than 1 year of oral contraceptive pill use, nulliparity, no breastfeeding, no tubal ligation, painful periods or endometriosis, polycystic ovarian syndrome or body mass index greater than 30 kg/m^2, and long-term talc use.

†Adjusted for study center, reference age, and study phase.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Hysterectomy Details and Cumulative Index of Experiences Among Women With Invasive Ovarian Cancer and Women in the Control Group Who Had Hysterectomy and Who Had No Personal History of Breast Cancer, Family History of Ovarian Cancer, or Early-Onset Breast

	No. of Women in the Case Group (n = 126)	No. of Women in the Control Group (n=156)	OR (95% CI)*	*P*
Age at hysterectomy (y)				
Younger than 35	35 (27.8)	36 (23.1)	1.00	
35–40	44 (34.9)	43 (27.6)	0.96 (0.50–1.82)	.89
41–46	30 (23.8)	39 (25.0)	0.77 (0.39–1.52)	.45
Older than 46	17 (13.5)	38 (24.4)	0.42 (0.20–0.90)	.02
P trend				.02
Time between hysterectomy and reference date (y)				
10 or less	27 (21.4)	42 (28.8)	1.00	
11–20	19 (15.1)	39 (25.0)	0.87 (0.39–1.92)	.72
21–30	45 (35.7)	43 (27.6)	1.84 (0.84–4.03)	.13
More than 30	35 (27.8)	29 (18.6)	2.17 (0.84–5.60)	.11
P trend				.04
Reason for hysterectomy				
Leiomyomas or heavy periods	78 (61.9)	89 (57.0)	1.00	
Endometriosis	10 (7.9)	13 (8.3)	0.92 (0.38–2.24)	.86
Prolapsed uterus	9 (7.1)	22 (14.1)	0.45 (0.19–1.05)	.06
Other	29 (23.0)	32 (20.5)	0.98 (0.54–1.77)	.94
Total number of risk factors†				
0–1	11 (8.7)	23 (14.7)	0.97 (0.39–2.38)	.94
2	21 (16.7)	41 (26.3)	1.00	
3	33 (26.2)	34 (21.8)	1.88 (0.92–3.86)	.08
4	33 (26.2)	35 (22.4)	1.83 (0.89–3.76)	.10
5 or more	28 (22.2)	23 (14.7)	2.45 (1.14–5.28)	.02
P trend				.01

OR, odds ratio; CI, confidence interval.

Data are n (%) unless otherwise specified.

*Adjusted for study center, reference age, and study phase.

†Risk factors include Jewish ethnicity, less than 1 year of oral contraceptive use, nulliparity, no breastfeeding, no tubal ligation, painful periods or endometriosis, polycystic ovarian syndrome or body mass index greater than 30 kg/m^2, and long-term talc use. The score was adjusted to estimate that which would have been observed before hysterectomy.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 4

Cumulative Risk of Developing Ovarian Cancer by Age 85 Years Using Oophorectomy-Adjusted Cumulative Incidence and the Relative Risks Associated With Each Level of the Risk-Factor Score

Total No. of Risk Factors	Probability of Developing Ovarian Cancer by Age 85 y Starting at Age								
	40	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**
0–1	1.2 (0.8–1.4)	1.2 (0.8–1.4)	1.1 (0.8–1.3)	1.1 (0.7–1.3)	1.0 (0.6–1.1)	0.8 (0.6–1.0)	0.7 (0.4–0.8)	0.5 (0.3–0.6)	0.2 (0.2–0.3)
2*	2.0	2.0	1.9	1.8	1.6	1.4	1.1	0.8	0.4
3	3.2 (2.6–4.2)	3.2 (2.6–4.2)	3.0 (2.5–4.0)	2.9 (2.3–3.8)	2.6 (2.1–3.4)	2.2 (1.8–2.9)	1.8 (1.4–2.3)	1.3 (1.0–1.7)	0.6 (0.5–0.8)
4	4.2 (3.2–5.4)	4.2 (3.2–5.4)	4.0 (3.0–5.1)	3.8 (2.9–4.9)	3.4 (2.6–4.3)	2.9 (2.2–3.8)	2.3 (1.8–3.0)	1.7 (1.3–2.2)	0.8 (0.6–1.1)
5 or more	6.6 (5.0–8.6)	6.6 (5.0–8.6)	6.3 (4.8–8.2)	5.9 (4.5–7.7)	5.3 (4.0–6.9)	4.6 (3.5–6.0)	3.6 (2.8–4.7)	2.6 (2.0–3.4)	1.3 (1.0–1.7)

* Two risk factors was chosen as the referent category.

Data are cumulative risk (95% confidence interval).

Exhibit F

CLINICAL OBSTETRICS AND GYNECOLOGY
Volume 55, Number 1, 3–23
© 2012, Lippincott Williams & Wilkins

Ovarian Cancer: Etiology, Risk Factors, and Epidemiology

JESSICA HUNN, MD and GUSTAVO C. RODRIGUEZ, MD
Division of Gynecologic Oncology, NorthShore University Healthsystem, Department of Obstetrics and Gynecology, University of Chicago, Evanston, Illinois

Abstract: Little is known regarding the early aspects of ovarian carcinogenesis. As a consequence, the identification of women at risk for the disease is based primarily on clinical grounds, with family history being the most important risk factor. In this review, we will discuss the various hypotheses regarding ovarian etiology and pathogenesis. In addition, we will discuss the epidemiology of ovarian cancer, including hereditary, reproductive, hormonal, inflammatory, dietary, surgical, and geographic factors that influence ovarian cancer risk.
Key words: ovarian cancer, epidemiology, risk factors, etiology, pathogenesis

Introduction

Epithelial ovarian cancer remains a highly lethal malignancy. It is the fourth to fifth leading cause of cancer deaths among women in the United States and causes more than 140,000 deaths annually in women worldwide. Despite intensive research efforts over the past decade directed toward improved detection and treatment of ovarian cancer, the majority of women diagnosed with ovarian cancer succumb to the disease. Progress in the fight against ovarian cancer has been hampered by a number of factors. These include late diagnosis, the absence of highly curative chemotherapy, and a high degree of molecular heterogeneity in ovarian tumors, a finding that is a direct consequence of the large tumor burden typical in most patients at the time of presentation. Despite the challenges, substantial progress has been made in our understanding of ovarian cancer biology, the potential mechanisms underlying protective factors, and our ability to identify women at increased risk of the disease. This is translating into more effective methods of prevention and treatment, and a corresponding fall in ovarian cancer incidence and mortality rates.[1]

Correspondence: Gustavo C. Rodriguez, MD, Division of Gynecologic Oncology, NorthShore University Healthsystem, 2650 Ridge Avenue, Evanston, IL. 60201 E-mail: grodriguez@northshore.org

The authors declare that they have nothing to disclose

Etiology

Because of the intra-abdominal location of the ovary as well as the preponderance

of advanced disease at presentation typical of most ovarian cancers, it has been difficult to characterize changes in the ovarian surface epithelium (OSE) consistent with intraepithelial neoplasia.[2] Thus, little is known regarding the very early molecular and genetic events associated with ovarian carcinogenesis. As a consequence, the etiology of ovarian cancer remains poorly understood, and even the cell of origin of epithelial ovarian cancer has not been conclusively defined. A common but unproven hypothesis is that ovarian cancers arise in OSE cell-lined inclusion cysts, which are nests of OSE that are entrapped in the ovarian stroma, and subjected to the stimulative influence of stromal growth factors. Evidence to support the OSE as the source of ovarian cancer includes: (1) the finding of activation of cancer preventive molecular pathways specifically in the OSE by the oral contraceptive pill (OCP), a known ovarian cancer preventive[3,4]; (2) description of premalignant, dysplastic changes in the OSE using classic pathologic criteria[5]; (3) colocalization of dysplastic histologic changes with either loss of tumor suppressor activity or overexpression of cyclooxygenase 2 in the OSE of high-risk ovaries[6,7]; and (4) the finding of a transition in some early ovarian cancers from a nonmalignant to malignant OSE.[8]

Recently, an alternative hypothesis has been proposed, which suggests that the cell of origin for ovarian cancer may involve cells that have originated in the fallopian tube.[9–13] This hypothesis is speculative, but supported by the finding that most ovarian cancers have a histology similar to that of the fallopian tube. In addition, fallopian tube cancer risk is markedly elevated in women with BRCA-related hereditary risk of ovarian cancer, and an unusually high incidence of histologic and molecular signatures associated with dysplasia have been identified in the fimbriated end of the fallopian tube in prophylactic oophorectomy specimens from women at high risk.[13,14] Further, careful examination of the fallopian tube in women with serous pelvic carcinoma has demonstrated a high incidence of endosalpinx involvement, or of coexistent tubal carcinomas, with similar alterations in p53 noted in the pelvic and fallopian tube lesions, suggesting that the lesions might be genetically related.[15,16] An unusually high incidence of p53 signatures has been noted even in the fimbriated ends of fallopian tubes removed for noncancerous indications in women at presumed population-based risk of ovarian cancer.[17] It is possible that the fimbriated end of the fallopian tube may be susceptible to neoplasia when exposed to dysplastic cells shed from the OSE or even in response to ovarian stromal factors released during ovulation.

PATHOGENESIS

It has been commonly believed that ovulation, with its associated disruption and subsequent repair of the ovarian epithelium, can lead to the acquisition of genetic damage in ovarian epithelial cells and, in turn, to ovarian cancer in susceptible individuals.[18–20] The "incessant ovulation" hypothesis for ovarian cancer is supported by a large volume of epidemiologic evidence linking ovulation with ovarian cancer risk[18,21–29] and by the finding that spontaneous ovarian cancers arise frequently in poultry hens, which ovulate daily.[30] Of note, alterations in p53 are common in epithelial ovarian cancer. In addition, in human as well as chicken ovarian adenocarcinomas, the incidence of p53 alterations correlates with the number of lifetime ovulatory events.[31] It is possible that ovulatory events predispose the ovarian epithelium to alterations in p53, leading to defective repair of DNA and thus ovarian cancer susceptibility. The mechanism(s) by which these changes could potentially lead to neoplastic transformation of the fallopian tube is unclear.

Case 3:16-md-02738-MAS-RLS Document 9732-4 Filed 05/07/19 Page 27 of 200 PageID: 33882

Under the incessant ovulation model, reproductive and hormonal factors such as OCP use and pregnancy have been presumed to alter ovarian cancer risk mainly through their inhibitory impact on ovulation. Although this hypothesis is attractive, it fails to explain completely the marked reduction in the degree of ovarian cancer risk associated with factors such as pregnancy and OCP use. For example, both of these factors confer a degree of ovarian cancer protection that is much greater than what would be expected simply based on the number of ovulatory cycles that are inhibited.[21,23] In addition, pregnancy is associated with a reduced risk of ovarian cancer even in women who are known to have ovulatory dysfunction and for whom the pregnant state has little impact on the number of lifetime ovulatory cycles.[32] Further, some studies have reported a relationship between increasing risk of epithelial ovarian cancer and increasing time since last birth.[33,34] These data support the hypothesis that reproductive and or hormonal factors impact ovarian cancer risk through additional biological mechanisms unrelated to ovulation inhibition.[35] Indeed, in addition to incessant ovulation, there is evidence in support of alternative hypotheses that have been proposed to explain ovarian cancer pathogenesis, including (1) the gonadotropin hypothesis, which purports that circulating gonadotropins stimulate the ovarian epithelium and promote neoplastic transformation,[36] (2) the hormonal hypothesis which suggests that reproductive hormones can interact directly with the ovarian epithelium to promote (estrogens and androgens) or protect against (progestins) carcinogenesis,[3,4,37] and (3) the inflammation hypothesis which argues that inflammatory mediators released either during ovulation or concomitant with disease processes such as endometriosis can damage the epithelium in the ovary and or fallopian tube.[38,39] Although none of these hypotheses can fully explain all ovarian cancers, it is likely that they all play a role, and that ovarian cancer pathogenesis is a multifactorial process, involving a complex interplay of biological events related to ovulation, inflammation, and gonadal/hormonal factors.

Risk Factors and Epidemiology

As a consequence of the fact that most ovarian cancers present in an advanced stage, the molecular or tissue biomarker changes associated with the very early aspects of ovarian epithelial carcinogenesis are not well known. Moreover, even if tissue biomarker changes predictive of neoplastic transformation of the OSE were known, the relative inaccessibility of the ovary would make it difficult to use this knowledge clinically to identify women at increased risk of the disease. In addition, despite extensive serum biomarker research, there is still a lack of robust serum biomarkers that can be used reliably to identify, in a timely way, the majority of women who are destined to develop ovarian cancer.[40] Thus, in contrast to other cancers such as that of the colon or cervix, there is insufficient tissue or other biomarker information to allow clinicians to identify women at risk, and risk identification is based primarily on epidemiologic factors (Table 1).

HEREDITARY

One of the most consistent and significant risk factors for ovarian cancer is a family history of ovarian cancer, particularly in first-degree relatives.[41,42] Schildkraut et al[43] examined the family histories of ovarian cases and controls who had been identified in conjunction with the Cancer and Steroid Hormone (CASH) Study in the early 1980s. The risks of ovarian cancer in first-degree and second-degree relatives of women with ovarian cancer were found to be increased 3.6- and 2.9-fold, respectively,

TABLE 1. Risk Factors for Epithelial Ovarian Cancer

Increased	Decreased	Indeterminate
Hereditary	Reproductive	Fertility drugs
Family history of ovarian cancer	Multiparity	Exercise
Personal history of breast cancer	Breastfeeding	Cigarette smoking
Alteration in BRCA1 or BRCA2		
Lynch syndrome	Hormonal	
	Oral contraceptives	
	Progestins	
Reproductive	Surgery	
Advanced age	Hysterectomy	
Nulligravity	Tubal ligation	
Infertility		
Hormonal		
Early age at menarche		
Late age at natural menopause		
Hormone replacement therapy		
Estrogen		
Androgens		
Inflammatory		
Perineal talc exposure		
Endometriosis		
Pelvic inflammatory disease		
Lifestyle		
Obesity		
Geography		
Extremes in latitude		

compared with women with no family history of ovarian cancer. Analysis of the CASH data also revealed that a family history of either breast or ovarian cancer increased the risk of both cancers in first-degree relatives.[43–45] The discovery of the *BRCA1* and *BRCA2* cancer susceptibility genes confirmed the hypothesis that a fraction of ovarian cancers arise in women with a genetic predisposition. It is now thought that about 10% to 12% of women with ovarian cancer carry germline mutations in the *BRCA1* or *BRCA2* genes.[46–50] An additional 2% to 3% are from families with hereditary nonpolyposis colorectal cancer (HNPCC) or Lynch syndrome. These families carry mutations in DNA repair genes and have as high as 10% to 13% lifetime risk of ovarian cancer, although colorectal, gastric, and endometrial cancers are more commonly seen.[51,52] Even among families with identical *BRCA1* or *BRCA2* mutations, there is heterogeneity with respect to the fraction of breast versus ovarian cancer that manifest and the age at onset. This suggests that genetic susceptibility is modified by other genetic or environmental factors. Cardinal features of hereditary cancer risk include a familial pattern suggestive of autosomal dominant inheritance, early onset, an excess of bilaterality (breast), multiple primaries (breast-ovary), and in the case of Lynch syndrome, an excess of cancers of the gastrointestinal and genitourinary tracts. Women with a familial pattern consistent with a significant risk of ovarian cancer should be referred for counseling and consideration of genetic testing (Table 2).[53]

BRCA

Families with *BRCA1* and *BRCA2* mutations represent the formerly separate syndromes of site-specific familial ovarian cancer and heredity breast/ovarian

TABLE 2. *Factors Suggestive of an Inherited Predisposition to Breast and/or Ovarian Cancer for Whom Referral for Genetic Evaluation Should Be Considered*

*BRCA**
- Personal history of both breast and ovarian cancer
- Personal history of ovarian cancer and a close relative with breast cancer at ≤ 50 y or ovarian cancer at any age
- History of ovarian cancer at any age combined with Ashkenazi Jewish ancestry
- History of breast cancer at ≤ 50 y and a close relative with ovarian or male breast cancer at any age
- Women of Ashkenazi Jewish ancestry and breast cancer at ≤ 40 y
- Women with a first-degree or second-degree relative with a known *BRCA1* or *BRCA2* mutation
- Women with bilateral breast cancer (particularly if the first cancer was at ≤ 50 y)
- Women with breast cancer at ≤ 50 y and a close relative with breast cancer at ≤ 50 y
- Women of Ashkenazi Jewish ancestry with breast cancer at ≤ 50 y
- Women with breast or ovarian cancer at any age and 2 or more close relatives with breast cancer at any age (particularly if at least 1 breast cancer was at ≤ 50 y)

Lynch
- Women with endometrial or colorectal cancer who have
  - At least 3 relatives with a Lynch/HNPCC-associated cancer (colorectal cancer, cancer of the endometrium, small bowel, ureter, or renal pelvis) in 1 lineage
    - One affected individual should be a first-degree relative of the other 2
    - At least 2 successive generations should be affected
    - At least 1 HNPCC-associated cancer should be diagnosed before age 50
- Women with synchronous or metachronous endometrial and colorectal cancer with the first cancer diagnosed before age 50

*Peritoneal and fallopian tube cancer should be considered as part of the spectrum of the hereditary breast/ovarian cancer syndrome.

HNPCC indicates hereditary nonpolyposis colorectal cancer.

Adapted from Schorge et al.[53] [Close relative is defined as a first, second, or third degree relative (ie, mother, sister, daughter, aunt, niece, grandmother, granddaughter, first cousin, great grandmother, great aunt)].

cancer.[54] Two thirds of these cancers are associated with alterations in *BRCA1* and the other third with alterations in *BRCA2*. The *BRCA* genes are tumor suppressor genes that play a role in the maintenance of genome integrity; they are involved in repair of double-strand DNA breaks, control of cell cycle checkpoint responses, and chromosomal segregation.[55] Affected individuals inherit an altered allele as well as normal wild-type allele for the *BRCA* genes. Loss of the wild-type alleles through either loss of heterozygosity or other somatic mutations in individuals with germline mutations in *BRCA1* and *BRCA2* leads to increases in genomic instability and tumorigenesis.[55]

The lifetime ovarian and breast cancer risks for women with *BRCA* mutations greatly surpasses that in the general population. Individuals from high-risk families with *BRCA1* mutations have an 87% cumulative risk of breast cancer by the age of 70. The lifetime risk of ovarian cancer in *BRCA1* mutation carriers is approximately 30% overall, but has been estimated to be as high as 44% in high-penetrance families.[56] The risk for breast and ovarian cancer is lower in women with mutations in *BRCA2*, with a 27% lifetime risk of ovarian cancer and an 84% risk of breast cancer.[57] Only a proportion of the women who carry *BRCA1* and *BRCA2* mutations develop ovarian cancer; the incomplete penetrance is thought to be due to multiple factors including the specific type and or location of the mutation, the status of modifying genes, epigenetic phenomena, and gene-environment interactions.[58,59] Of note, the estimated frequency of *BRCA* mutations in the general population is relatively low (1 in 300 to 1 in 800 individuals in the United States), but is considerably higher in those of Ashkenazi Jewish heritage (1 in 50).[60] Thus, in women with breast or ovarian cancer, those of Ashkenazi Jewish heritage are significantly more likely to harbor an alteration in *BRCA1* or *BRCA2*.

Lynch Syndrome (HNPCC)

A strong family history of early onset colon or endometrial cancer, or multiple malignancies of the gastrointestinal and genitourinary system should alert clinicians to the possibility of Lynch syndrome.[53] In addition to a significant lifetime risk of developing colon cancer, HNPCC patients have an increased risk of ovarian (12%) and endometrial cancers (40% to 60%).[61] These patients carry a mutation in the DNA mismatch repair genes MSH2, MLH1, PMS1, and PMS2, leading to genomic instability and cancer risk.[62] Similar to *BRCA*-related cancers, it has been observed that women with Lynch syndrome develop ovarian cancer at a younger age than women with sporadic ovarian cancer, with a mean age of 48. In half of the cases, ovarian and or endometrial cancers occur as many as 5 or more years before the onset of colon cancer, thereby being the sentinel event alerting clinicians to the possible risk of HNPCC.[63] Patients who have developed malignancies suspicious for Lynch syndrome often undergo genetic assessment in a stepwise fashion starting with screening of tumor (uterus or colon) for mismatch repair defects.[53] Patients with abnormalities on immunohistochemical evaluation of MLH1, MSH2, MSH6, and PMS2 protein expression or microsatellite instability will then typically undergo full sequence analysis of relevant genes as directed by immunohistochemical results.

REPRODUCTIVE

Parity

Case-control evidence has consistently shown that pregnancy lowers ovarian cancer risk. One pregnancy lowers ovarian cancer risk by as much as one third and the reduction in risk increases with each additional pregnancy.[21,23–27] The protective effect lingers for as long as 1 to 2 decades, but then wanes with increasing time since last birth.[33,34] In addition, pregnancy at a later age is more protective than pregnancy early in life. In fact, a pregnancy after the age of 35 is twice as protective against ovarian cancer as a pregnancy before the age of 25. It has been proposed that this would suggest a protective effect of pregnancy that is unrelated to effects on ovulation, and supporting the notion that pregnancy may clear premalignant or damaged cells from the ovary.[64–65] Infertility is associated with a 2-fold increased relative risk (RR) of ovarian cancer. Data on the impact of fertility drug use on risk have been inconsistent, perhaps because of the confounding influences of infertility and pregnancy on ovarian cancer risk.[66–69] Of note, similar to women who are fertile, women treated for infertility who successfully achieve a live birth benefit from a reduction in ovarian cancer risk.

OCP Use

Numerous case-control studies have shown that OCP use is associated with a decreased risk of ovarian cancer.[21,70] Three or more years of OCP use reduces the risk of developing epithelial ovarian cancer by 30% to 50%.[22,71] The association increases with the duration of use and appears to be independent of inherent ovarian cancer risk.[23,72] Furthermore, the duration of protection effect lasts for more than 10 to 20 years after the last use. These data are quite similar to the epidemiologic data related to parity, suggesting that parity and OCP use may share a common biological mechanism underlying their ovarian cancer protective effect.

Breastfeeding

Although the results of published studies are inconsistent, the weight of the published evidence suggests that breastfeeding lowers ovarian cancer risk. Danforth evaluated the impact of breastfeeding on ovarian cancer risk in a large study of 391

ovarian cancer cases and over 149,000 total participants.[73] Analysis was confined to parous women to evaluate the impact of breastfeeding independent of parity. The median duration of breastfeeding among women who breastfed was 9 months. As compared with never breastfeeding, any breastfeeding was not associated with a statistically significant reduction in ovarian cancer risk. However, among those women who breastfed for 18 months or more, a significant 34% decrease in ovarian cancer risk was noted as compared with never breastfeeding. A similar protective effect of breastfeeding was noted in a case-control study of parous women in New Hampshire, but only for women who had either breastfed all children, or the last born child.[74] No protective effect was found when the last born child was not breastfed. The authors speculated that breastfeeding may "reset pregnancy-related influences on ovarian cancer risk." In contrast, Jordan found a modest 2% reduction in ovarian cancer risk associated with breastfeeding, and no additional benefit from individual lactation episodes >12 months. In addition, the protective effect did not hold for serous borderline or mucinous subtypes, but was generally maintained for other histologic subtypes of ovarian cancer.[75]

HORMONAL

There is mounting evidence that the ovarian epithelium is a hormonally responsive target organ whose biology can be impacted strongly by the local hormonal environment. The normal ovarian epithelium expresses receptors for most members of the steroid hormone superfamily, including estrogens, progestins, retinoids, vitamin D, and androgens. In addition, the ovarian epithelium contains gonadotropin receptors and nonhormonal targets such as the cyclooxygenase pathway. There is therefore the potential for reproductive and environmental factors to have an impact on ovarian cancer risk through a direct biological interaction of hormonal and nonhormonal agents on the ovarian epithelium. Recent studies have indeed shown that reproductive hormones can have potent biological effects directly on the ovarian epithelium, thus impacting ovarian cancer risk. Progestins, for example, have been shown to induce apoptosis, one of the most important molecular pathways in vivo for the prevention of cancer and a pathway that mediates the action of many known chemopreventive agents. It has been proposed that progestin-mediated apoptotic effects may be a major mechanism underlying the ovarian cancer protective effects of OCP use and pregnancy (a high progestin state). Similarly, retinoids, vitamin D, and nonsteroidal anti-inflammatory drugs may have biological effects on the ovarian epithelium that are cancer preventive, whereas estrogens and androgens may have stimulatory effects on the ovarian epithelium, leading to an increased ovarian cancer risk.[3,4,37,76]

Gonadotropins

As early at the 1980s, Cramer proposed the gonadotropin hypothesis as a potential mechanism underlying ovarian carcinogenesis.[24] He proposed that elevated circulating levels of gonadotropins related to either the menopause or ovulatory events might stimulate the OSE and promote neoplastic transformation. The biological mechanisms underlying the gonadotropin hypothesis have not been well characterized, however, and the theory has fallen short in fully explaining the impact of hormonal and reproductive events on ovarian cancer risk. Recently, an excellent review by Choi has summarized the evidence in support of or against the gonadotropin hypothesis, and the published data have generally yielded inconsistent findings.[77] For example, although gonadotropin receptors have been shown to be expressed in the normal

ovarian epithelium and ovarian neoplasms, an association between serum levels of gonadotropins and ovarian cancer has not been conclusively established. Similarly, the known reduction in ovarian cancer risk associated with pregnancy and OCP use, conditions where gonadotropins are suppressed, supports the gonadotropin hypothesis; yet hormone replacement therapy, which also suppresses gonadotropins, is associated with an increase in ovarian cancer risk. Finally, gonadotropins have been shown to both inhibit and stimulate carcinogenesis in vitro, and animal data have been similarly inconsistent.

Progestins

The biological mechanism underlying the protective effect of OCP use has historically been presumed to be related to the inhibitory effect of OCPs on ovulation, and, in turn, to a lessening in the extent of ovulation-induced genetic damage accumulated in the OSE. Recent animal data, however, suggest that the OCP may have a profound, direct chemopreventive effect in the OSE, mediated by the progestin component. A 3-year study in primates has demonstrated that the progestin component of an OCP has a potent apoptotic effect on the ovarian epithelium, providing support for the hypothesis that OCPs may lower ovarian cancer risk through progestin induction of cancer preventive molecular pathways in the ovarian epithelium.[3,4] The apoptosis pathway is arguably one of the most important in vivo mechanisms for cancer prevention. Activation of apoptosis leads to the efficient disposal of cells that have undergone irreparable genetic damage and that are prone to neoplastic transformation.[78] It is thus a key molecular pathway for the elimination of premalignant cells in vivo. It is a biological mechanism associated with many known chemopreventive agents,[79–86] and pharmacologic agents that selectively enhance apoptosis have been shown to lower the risk of a variety of cancers in animals and in humans.[87] In addition, in both animal models of cancer as well as in humans, the efficacy of cancer preventive agents has been shown to correlate with the degree of apoptosis induced.[87–90] Conversely, mutations in the genes involved in the apoptosis pathway have been shown to be associated with enhanced cancer risk.[91] The finding that progestins activate this critical pathway in the ovarian epithelium raises the possibility that progestin-mediated apoptotic effects, and not solely ovulation inhibition as has been previously assumed, may underlie the reduction in ovarian cancer risk associated with routine OCP use and pregnancy.

A growing body of published human data is supportive of the notion that a biological effect related to progestins may be a major mechanism underlying the cancer preventive effect for both the OCP as well as pregnancy, which confers potent protection against subsequent ovarian cancer and which is associated with high serum progesterone levels:

(a) An analysis of the data from the CASH, has demonstrated that use of progestin-potent OCPs confers greater protection against ovarian cancer than use of OCPs containing weak progestin formulations.[92]
(b) Further support for progestins as ovarian cancer preventives has come from an analysis of data from the WHO by Risch, demonstrating a 60% reduction in the risk of nonmucinous ovarian cancer in women who have ever used Depo-medroxyprogesterone acetate, a progestin-only contraceptive.[37] Progestin-only contraceptives do not reliably inhibit ovulation. Thus, the 60% reduction in ovarian cancer risk from a progestin-only contraceptive is further evidence that progestins have a direct chemopreventive effect on the ovary.
(c) In addition, epidemiologic evidence has suggested that twin pregnancy may be more protective against

subsequent ovarian cancer than singleton pregnancy. Previously, it was presumed that women who have twins would be at greater risk of ovarian cancer, presumably due to an increased likelihood of more lifetime ovulatory events as compared with women who do not have twins, and the notion that increased ovulation would confer greater risk of ovarian epithelial damage. Because women with twin pregnancy have higher progesterone levels than women with singleton pregnancy, it has been proposed that the data regarding twin pregnancy are supportive of the notion of a biological effect of progesterone as conferring ovarian cancer protection, and that the effect is dose dependent.[64]

(d) Finally, pregnancy at a later age is more protective than pregnancy early in life, and pregnancy after the age of 35 is twice as protective against ovarian cancer as a pregnancy before the age of 25. It has been proposed that this would suggest a protective effect of pregnancy that is unrelated to effects on ovulation, and supporting the notion that pregnancy may clear premalignant or damaged cells from the ovary.[64,65] Reproductive factors such as pregnancy and OCP use may thus impact ovarian cancer risk not only through inhibition of ovulation, but also through a progestin-mediated chemopreventive effect that clears genetically damaged cells from the ovarian epithelium.

Estrogens

Data regarding the impact of estrogens on ovarian cancer risk are mainly derived from case-control series examining the impact of OCP use or hormone replacement therapy on ovarian cancer risk. As discussed above, use of estrogen/progestin combination OCPs has been shown to consistently lessen ovarian cancer risk.[71] Of note, however, in primates receiving OCPs, estrogens have been shown to partly abrogate the effect of progestins on chemopreventive endpoints such as apoptosis in the OSE, suggesting that estrogens may counteract the cancer preventive effect of progestins.[3,4] Published evidence in postmenopausal women would support this conclusion. Several large case-control studies suggest that estrogen replacement therapy increases ovarian cancer risk 2-fold, and that the addition of progestins to hormone replacement therapy partly neutralizes this enhanced risk.[93–97] Whether or not estrogen replacement therapy increases the risk for all ovarian cancers, or selectively promotes the development of specific histologic subtypes of ovarian cancer is unclear. For example, an increase in risk for endometrioid ovarian tumors has been reported among women who have used postmenopausal estrogen replacement.[97,98] A more recent study, however, has shown that menopausal hormone replacement use conferred an increased risk for all histologic subtypes of ovarian cancer except for mucinous, where risk was reduced.[99]

Androgens

It has been proposed that androgens may be associated with increased ovarian cancer risk, but the evidence is not conclusive.[37,100] Data in support of a link between androgens and ovarian cancer risk include: (1) androgen receptors (ARs) are expressed in the OSE, thereby providing a means by which androgens can have a direct biological effect in the organ; (2) most ovarian cancers express AR, and antiandrogens inhibit ovarian cancer growth; (3) oral contraceptives, potent ovarian cancer preventives, significantly lower ovarian androgen production; (4) ovarian cancer risk is increased in conditions such as polycystic ovary syndrome, which is associated with elevated serum androgen levels; (5) use of androgenic agents such as testosterone or danazol may increase ovarian cancer risk.[101,102] In contrast, however, increased

activity of the AR gene may inhibit ovarian carcinogenesis. In addition, a recent case-control study evaluating clinical surrogates for an androgenic milieu such as a history of polycystic ovary syndrome, acne or hirsutism failed to demonstrate that androgen excess is associated with increased ovarian cancer risk.[101] Finally, use of androgenic OCPs does not increase ovarian cancer risk as compared with nonandrogenic OCPs.[103]

INFLAMMATION

Ness was the first to propose that inflammatory factors might be involved in ovarian carcinogenesis.[104] In her comprehensive review in 1999, she noted that the incessant ovulation and gonadotropin hypotheses failed to adequately explain the enhanced risk of ovarian cancer associated with talc use, endometriosis and pelvic inflammatory disease (PID), as well as the protective effects associated with hysterectomy and tubal ligation. A growing body of evidence suggests that the ovarian epithelium and fallopian tube are exposed chronically to an inflammatory milieu related to the normal functions of ovulation and menstruation.[105] Pro-inflammatory cytokines are present in ovulatory fluid and also in menstrual effluent that comes into contact with the fallopian tube. These same cytokines are markedly elevated in epithelial ovarian cancers. In addition, inflammatory mediators are markedly increased in disease states such as endometriosis and PID. Recently, elevated serum levels of C-reactive protein have been shown to be associated with an increased subsequent risk of ovarian cancer.[106,107] In addition, in a prospective case-control study of 230 women with ovarian cancer and 432 individually matched controls nested within three prospective cohorts, prediagnostic circulating levels of inflammatory cytokines, such as the interleukins, have been shown to be elevated in women who subsequently developed ovarian cancer. These data provide more direct evidence that inflammation may be associated with ovarian cancer risk.[108] Interestingly, OCPs, which as described above, markedly lower ovarian cancer risk, confer a number of biological effects that can mitigate inflammatory influences in the genital tract, including inhibiting ovulation, lowering the risk of PID, and reversing endometriosis.[109]

Talc

Evidence demonstrating an association between talc use and an increased risk of ovarian cancer suggests that environmental toxins can enter the lower genital tract and migrate upward through the uterus and fallopian tubes to enter the peritoneal cavity and act as ovarian carcinogens. Talcum powder was first implicated in the risk of ovarian cancer in the 1960s when it was found to be biologically similar to asbestos which is a known carcinogen. Subsequent studies in animals and humans demonstrated not only that talc deposited in the gynecologic tract could reach the ovaries, but also the finding of talc particles in ovarian neoplasms.[110] Subsequent case-control studies of talc use and risk of ovarian cancer have shown a strong association, including a meta-analysis of 16 studies that included 11,933 women demonstrating a 33% increased risk of ovarian cancer.[111–115]

Endometriosis

Endometriosis has been consistently shown to be associated with an increased risk of ovarian cancer, with odds ratios of approximately two.[104,116] The underlying mechanism is not fully characterized. It has been proposed that chronic inflammation can lead to neoplastic transformation of endometriotic implants. In addition, it is possible that the endometriotic state leads to a relative progesterone “resistance”, thereby mitigating the potential protective effects of the hormone.[117,118] The most common histologic subtypes of ovarian cancer associated

with endometriosis are clear cell and endometrioid carcinomas.[119]

PID

PID occurs as predominantly a consequence of sexually transmitted diseases and manifests clinically as a marked inflammatory process involving the uterus, fallopian tubes, and ovaries. Limited case-control evidence suggests an increased risk of ovarian cancer among women who have had PID.[120,121] The association appears to be most pronounced in women who have had PID at a young age, or who are infertile, which is also an ovarian risk factor. In the largest study to date, with over 67,000 women with PID and over 135,000 controls, the adjusted hazard ratio for ovarian cancer in women with PID was 1.92, increasing to 2.46 in women who had had 5 or more episodes of PID. The adjusted hazard ratio was higher for women aged 35 or younger.[121]

SURGERY

Hysterectomy and tubal ligation are associated with a reduction in the risk of developing ovarian cancer. In a meta-analysis of 12 case-control studies, hysterectomy (without oophorectomy or salpingectomy) was associated with a 34% reduction in the risk of ovarian cancer.[29] Women who underwent a tubal ligation also had a 34% risk reduction compared with women who did not.[122] The protective effect of surgery also extends to women at hereditary risk of ovarian cancer. A case-control study by the Hereditary Ovarian Cancer Clinical Study Group has shown that tubal ligation lowered the rate of ovarian cancer in women with *BRCA1* alterations by 60%.[123] The combination of tubal ligation and OCP use reduced the risk even further. Of note, no protective effect of tubal ligation was seen among carriers of the *BRCA2* mutation. The mechanism for the protective effect of tubal ligation and hysterectomy is not known, but theoretically could be explained by blockage of access of environmental carcinogens to the ovaries. Another proposed mechanism is that surgery to remove uterus or fallopian tubes may affect the ovarian circulation or plasma hormone levels in ways that lower ovarian cancer risk.[124] Finally, if the fallopian tube is indeed the source of some ovarian cancers, then removing some of the tube may be expected to lower cancer risk.

LIFESTYLE

Obesity

It is likely that obesity increases the risk of ovarian cancer, but the degree of effect is modest. A systematic review reported a small association between body mass index (BMI) >30 and ovarian cancer risk with an odds ratio of 1.3 [95% confidence interval (CI), 1.1-1.5].[125] In the Cancer Prevention Study, a prospective cohort study of 495,477 women followed for 16 years, a relationship was noted between high BMI and ovarian cancer mortality.[126] The RR of death from ovarian cancer among women with a BMI of 35 to 40 was 1.51 compared with those of normal weight. Findings from the Nurses' Health Study indicated a 2-fold increased risk of premenopausal ovarian cancer associated with a high BMI.[127] In addition, a meta-analysis showed an association between obesity and ovarian cancer with a 40% increase in risk in the heaviest versus the lightest women in population-based case-control studies.[128] A recent study by Leitzman prospectively followed 94,525 patients over a 7-year period.[129] Overall, the women with a BMI > 30 were 1.26 times more likely to have developed ovarian cancer, though those findings were not statistically significant. Among a subgroup of women who had never used hormone replacement therapy, the women who were obese were 1.83 times more likely to develop ovarian cancer. In

Case 3:16-md-02738-MAS-RLS Document 9732-4 Filed 05/07/19 Page 36 of 200 PageID: 33891

women who had used hormone replacement therapy, there was no association between obesity and ovarian cancer. The authors speculated that obesity is associated with enhanced ovarian cancer risk through a hormonal mechanism. Obesity is known to increase adrenal secretion of androgens, and is generally associated with an increased endogenous production of estrogens.[130]

Diet

Numerous studies have attempted to identify dietary factors that may influence ovarian cancer risk. Overall, the results have been inconsistent or conflicting. The balance of the evidence has failed to conclusively demonstrate that consumption of any macronutrient or micronutrient significantly alters ovarian cancer risk. A case-control study in Italy comparing 455 cases with ovarian cancer to 1385 age-matched controls revealed an increased RR for ovarian cancer associated with meat consumption of >7 portions versus less than 4 portions per week (RR 1.6; 95% CI, 1.2-2.12) and butter versus fat consumption (RR 1.9; 95% CI, 1.20-3.11). Dietary risk factors that decreased risk included whole-grain bread and pasta consumption.[131] A larger prospective cohort study of 29,083 women in the United States found that egg consumption of 2 to 4 times per week as well as increased intake of carbohydrates and dairy increased the RR of developing ovarian cancer, whereas consumption of green leafy vegetables significantly decreased risk (RR 0.44, 95% CI, 0.25-0.79), but there was no association with dietary fat, as well as intake of meats, breads cereals, and starches and ovarian cancer risk.[132]

Studies evaluating the intake of specific foods or food groups on the subsequent development of ovarian cancer have similarly yielded inconsistent results. In one study, protective foods included olive and vegetable oils, fish, peas, beans, and lentils.[133] Vegetable consumption was found to be protective in one study[134] but another study that examined the effect of consumption of vegetables and fruits noted no benefit.[135] In another large study, risk of ovarian cancer was studied after consumption of fruit and vegetables. There was no association found between high consumption of fruits and vegetables and ovarian cancer risk.[136] A study in 2006 suggested that milk and milk products are associated with an increased ovarian cancer risk.[137] However, the Netherlands Cohort Study on Diet and Cancer which followed 62,573 women for 11.3 years and included 252 cases with ovarian cancer found no association between lactose and dairy intakes and the development of ovarian cancer.[138]

In attempt to further clarify dietary associations with ovarian cancer risk, 2 studies evaluated general dietary patterns as opposed to specific foods. Overall diet was evaluated in the prospective California Teachers Study.[139] A total of 97,292 women completed a baseline dietary assessment of which 311 developed epithelial ovarian cancer. Five major dietary patterns were compared: (1) plant-based; (2) high protein/high fat; (3) high carbohydrate; (4) ethnic; (5) salad and wine. Although women who followed a plant-based diet had a slightly higher risk of ovarian cancer (RR 1.65, 95% CI, 1.07-2.54), the authors concluded that their results did not show convincing associations between dietary patterns and ovarian cancer risk. A recent study published in 2011 evaluated the association between a Healthy Eating Index and ovarian cancer.[140] The Healthy Eating Index reflects adherence to current USDA dietary Guideline for Americans. This population-based case-control study had a total of 205 women with ovarian cancer and 390 controls. Based on their results, the authors concluded that neither individual food groups nor dietary quality showed potential for preventing ovarian cancer.

Exercise

There is no firm relationship between exercise and ovarian cancer risk. Studies to date are small and generally inconclusive, with results ranging from suggesting no association, to a finding of a modest benefit from exercise, to even a possible adverse effect of vigorous exercise on ovarian cancer risk.[141–144] Pan et al[145] examined survey responses from over 400 women with ovarian cancer and over 2100 healthy women from The Canadian National Enhanced Cancer Surveillance System. Women who reported moderate levels of recreational physical activity or who held jobs with moderate or strenuous physical activity had a reduced risk of ovarian cancer with an odds ratio of 0.67 (0.50 to 0.88). A large study from the Netherlands Cohort Study consisting of 62,573 women who were surveyed regarding their physical activity yielded similar conclusions. Two hundred fifty-two cases of ovarian cancer were identified after 11.3 years of follow-up. Compared with women who exercised <30 minutes per day, women who spent >60 minute per day in moderate exercise had a RR of 0.78 for the development of ovarian cancer. Women who spent >2 hours per week on recreational biking and walking had an even lower risk (RR 0.65; 95% CI, 0.41-1.01) compared with women who did no exercise.[146] In contrast, in the very large Nurses Health Study, although moderate activity was found to be protective against subsequent ovarian cancer, frequent vigorous exercise was associated with increased risk.[143] The underlying mechanism(s) potentially mediating the effects of exercise on ovarian risk are not well known. Hormonal changes associated with physical activity can cause anovulation and decrease the risk of obesity thereby lowering estrogens and risk, but possibly increase gonadotropins which may increase risk.

Cigarette Smoking

The effect of smoking on ovarian cancer risk has not been well defined. The most intriguing finding has been an association between current or past smoking and an increase in mucinous ovarian cancer, although the association does not apply to other histologic subtypes.[147–151] The biological basis underlying any association between smoking and ovarian cancer is not well understood. Nicotine and its metabolites have been identified in ovarian tissue.[152] Thus, it is plausible that these agents can cause direct DNA damage in the OSE. In addition, cigarette smokers have been found to have higher circulating levels of gonadotropins and androgens, both of which can have adverse effects on risk. On the other hand, smokers may have earlier onset of menopause which would be expected to lower risk.[153–155]

GEOGRAPHY

Worldwide, there is a geographic distribution for ovarian cancer, with increasing incidence commensurate with latitudinal distance from the equator.[156] The same pattern holds in the United States where there is a significant north-south gradient, favoring a higher ovarian cancer risk in northern versus southern latitudes in the United States. Lefkowitz has correlated population-based data regarding ovarian cancer mortality in large cities across the United States with geographically based long-term sunlight data reported by the National Oceanic and Atmospheric Administration, demonstrating a statistically significant inverse correlation between regional sunlight exposure and ovarian cancer mortality risk.[157] Given that sunlight induces production of previtamin D in the skin, it is interesting to speculate that vitamin D might confer protection against ovarian cancer by direct biological effects in the nonmalignant ovarian epithelium, similar to that induced by progestins. For example through induction of apoptosis and/or transforming growth factor-β in the ovarian epithelium,

vitamin D may cause the selective removal of nonmalignant, but genetically damaged ovarian epithelial cells.[158,159] A small case-control study supports the notion that vitamin D confers ovarian cancer prevention, at dosages of vitamin D easy to achieve through the diet. As compared with a low dietary intake of vitamin D, a high dietary intake of vitamin D was associated with a 50% reduction in ovarian cancer risk.[160]

References

1. American Cancer Society. *Cancer Facts & Figures 2011*. Atlanta: American Cancer Society; 2011.
2. Brewer MA, Johnson K, Follen M, et al. Prevention of ovarian cancer: intraepithelial neoplasia. *Clin Cancer Res*. 2003; 9:20–30.
3. Rodriguez GC, Walmer DK, Cline M, et al. Effect of progestin on the ovarian epithelium of Macaques: cancer prevention through apoptosis? *J Soc Gynecol Invest*. 1998;5:271–276.
4. Rodriguez GC, Nagarsheth NP, Lee KL, et al. Progestin-induced apoptosis in the macaque ovarian epithelium: differential regulation of transforming growth factor-beta. *J Natl Cancer Inst*. 2002;94:50–60.
5. Scully RE. Pathology of ovarian cancer precursors. *J Cell Biochem*. 1995;23 (suppl):208–218.
6. Roland IH, Yang WL, Yang DH, et al. Loss of surface and cyst epithelial basement membranes and preneoplastic morphologic changes in prophylactic oophorectomies. *Cancer*. 2003;98: 2607–2623.
7. Yang DH, Smith ER, Cohen C, et al. Molecular events associated with dysplastic morphologic transformation and initiation of ovarian tumorigenicity. *Cancer*. 2002;94:2380–2392.
8. Plaxe SC, Deligdisch L, Dottino PR, et al. Ovarian intraepithelial neoplasia demonstrated in patients with stage I ovarian carcinoma. *Gynecol Oncol*. 1990;38:367–372.
9. Medeiros F, Muto M, Lee Y, et al. The tubal fimbria is a preferred site for early adenocarcinoma in women with familial ovarian cancer syndrome. *Am J Surg Pathol*. 2006;30:230–236.
10. Lee Y, Medeiros F, Kindelberger D, et al. Advances in the recognition of tubal intraepithelial carcinoma: applications to cancer screening and the pathogenesis of ovarian cancer. *Adv Anat Pathol*. 2006;13:1–7.
11. Piek JM, Kenemans P, Verheijen RH, et al. Intraperitoneal serous adenocarcinoma: a critical appraisal of three hypotheses on its cause. *Am J Obstet Gynecol*. 2004;191:718–732.
12. Medeiros F, Muto M, Lee Y, et al. The tubal fimbria is a preferred site for early adenocarcinoma in women with familial ovarian cancer syndrome. *Am J Surg Pathol*. 2006;30:230–236.
13. Crum CP, Drapkin R, Kindelberger D, et al. Lessons from BRCA: the tubal fimbria emerges as an origin for pelvic serous cancer. *Clin Med Res*. 2007;5: 35–44.
14. Roh MH, Yassin Y, Miron A, High-grade fimbrial-ovarian carcinomas are unified by altered p53, PTEN and PAX2 expression. Mod Pathol. 2010;23: 1316–1324.
15. Kindelberger DW, Lee Y, Miron A, et al. Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: evidence for a causal relationship. *Am J Surg Pathol*. 2007;31:161–169.
16. Chivukula M, Niemeier LA, Edwards R, Carcinomas of distal fallopian tube and their association with tubal intraepithelial carcinoma: do they share a common "precursor" lesion? Loss of heterozygosity and immunohistochemical analysis using PAX 2, WT-1, and P53 markers. Obstet Gynecol. 2011; 2011:858647.
17. Shaw PA, Rouzbahman M, Pizer ES, et al. Candidate serous cancer precursors in fallopian tube epithelium of BRCA1/2 mutation carriers. *Mod Pathol*. 2009;22:1133–1138.
18. Schildkraut JM, Bastos E, Berchuck A. Relationship between lifetime ovulatory cycles and overexpression of mutant p53

Case 3:16-md-02738-MAS-RLS Document 9732-4 Filed 05/07/19 Page 39 of 200 PageID: 33894

in epithelial ovarian cancer. *J Natl Cancer Inst*. 1997;89:932–938.
19. Ames BN, Gold LS. Too many rodent carcinogens: mitogenesis increases mutagenesis. *Science*. 1990;249:970–971.
20. Preston-Martin S, Pike MC, Ross RK, et al. Increased cell division as a cause of human cancer. *Cancer Res*. 1990;50: 7415–7421.
21. Whittemore AS, Harris R, Itnyre J. Characteristics relating to ovarian cancer risk: collaborative analysis of 12 US case-control studies. IV. The pathogenesis of epithelial ovarian cancer. Collaborative Ovarian Cancer Group. *Am J Epidemiol*. 1992;136:1212–1220.
22. Anonymous. The reduction in risk of ovarian cancer associated with oral-contraceptive use. The Cancer and Steroid Hormone Study of the Centers for Disease Control and the National Institute of Child Health and Human Development. *N Engl J Med*. 1987;316:650–655.
23. Risch HA, Weiss NS, Lyon JL, et al. Events of reproductive life and the incidence of epithelial ovarian cancer. *Am J Epidemiol*. 1983;117:128–139.
24. Cramer DW, Hutchinson GB, Welch WR, et al. Determinants of ovarian cancer risk. I. Reproductive experiences and family history. *J Natl Cancer Inst*. 1983;71:711–716.
25. Joly DJ, Lilienfeld AM, Diamond EL, et al. An epidemiologic study of the relationship of reproductive experience to cancer of the ovary. *Am J Epidemiol*. 1974;99:190–209.
26. Hildreth NG, Kelsey JL, LiVolsi VA, et al. An epidemiologic study of epithelial carcinoma of the ovary. *Am J Epidemiol*. 1981;114:398–405.
27. Franceschi S, La Vecchia C, Helmrich SP, et al. Risk factors for epithelial ovarian cancer in Italy. *Am J Epidemiol*. 1982;115:714–719.
28. Rosenberg L, Shapiro S, Slone D, et al. Epithelial ovarian cancer and combination oral contraceptives. *JAMA*. 1982; 247:3210–3212.
29. Whittemore AS, Harris R, Itnyre J, et al. Characteristics relating to ovarian cancer risk: collaborative analysis of 12 US case-control studies I. Methods. Collaborative Ovarian Cancer Group. *Am J Epidemiol*. l992;136:1175–1183.
30. Fredrickson TN. Ovarian tumors of the hen. *Environ Health Perspect*. 1987; 73:35–51.
31. Hakim AA, Barry CP, Barnes HJ, et al. Ovarian adenocarcinomas in the laying hen & women have similar alterations in p53, ras, & HER-2/neu. *Cancer Prev Res*. 2009;2:114–121.
32. Mosgaard BJ, Lidegaard O, Andersen AN. The impact of parity, infertility and treatment with fertility drugs on the risk of ovarian cancer. *Acta Obstel Gynecol Scand*. 1997;76:89–95.
33. Cooper GS, Schildkraut JM, Whittemore AS, et al. Pregnancy recency and risk of ovarian cancer. *Cancer Causes Control*. 1999;10:397–402.
34. Albrektsen G, Heuch I, Kvale G. Reproductive factors and incidence of epithelial ovarian cancer: a Norwegian prospective study. *Cancer Causes Control*. 1996;7:421–427.
35. Adami HO, Hsieh CC, Lambe M, et al. Parity, age at first childbirth, and risk of ovarian cancer. *Lancet*. 1994;344: 1250–1254.
36. Cramer DW, Welch WR. Determinants of ovarian cancer risk. II. Inferences regarding pathogenesis. *J Natl Cancer Inst*. 1983;71:717–721.
37. Risch HA. Hormonal etiology of epithelial ovarian cancer, with a hypothesis concerning the role of androgens and progesterone. *J Natl Cancer Inst*. 1998; 90:1774–1786.
38. Ness RB, Cottreau C. Possible role of ovarian epithelial inflammation in ovarian cancer. *J Natl Cancer Inst*. 1999;91: 1459–1467.
39. Shan W, Liu J. Inflammation: a hidden path to breaking the spell of ovarian cancer. *Cell Cycle*. 2009;8: 3107–3111.
40. Cramer DW, Bast RC Jr, Berg CD, et al. Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. *Cancer Prev Res*. 2011;4:365–374.
41. Wu ML, Whittemore AS, Paffenbarger RS Jr, et al. Personal and environmental characteristics related to epithelial ovarian

cancer. I. Reproductive and menstrual events and oral contraceptive use. *Am J Epidemiol*. 1988;128:456–466.

42. Coch M, Gaedke H, Jenkins H. Family history of ovarian cancer patients: a case-controlled study. *Int J Epidemiol*. 1989;18:782–785.
43. Schildkraut JM, Risch N, Thompson WD. Evaluating genetic association among ovarian, breast, and endometrial cancer: evidence for a breast/ovarian cancer relationship. *Am J Human Genet*. 1989;4:521–529.
44. Claus EB, Risch N, Thompson WD. Genetic analysis of breast cancer and steroid hormone study. *Am J Hum Genet*. 1991;48:232–242.
45. Claus EB, Risch N, Thompson WD. The calculation of breast cancer risk for women with a first-degree family history of ovarian cancer. *Breast Cancer Res Treat*. 1993;28:115–120.
46. Berchuck A, Cirisano F, Lancaster JM, et al. Role of BRCA1 mutation screening in the management of familial ovarian cancer. *Am J Obstet Gynecol*. 1996; 175:738–746.
47. Sinilnikova OM, Mazoyer S, Bonnardel C, et al. Gynecologic manifestations of hereditary nonpolyposis colorectal cancer. From inherited to sporadic disease. *Fam Cancer*. 2006;5:15–20.
48. Sogaard M, Kjaer SK, Gayther S. Ovarian cancer and genetic susceptibility in relation to the BRCA1 and BRCA2 genes. Occurrence, clinical importance and intervention. *Acta Obstet Gynecol Scand*. 2006;85:93–105.
49. Schwartz MD, Peshkin BN, Tercyak KP, et al. Decision making and decision support for hereditary breast-ovarian cancer susceptibility. *Health Psychol*. 2005;24 (4, suppl):S78–S84.
50. Nelson HD, Huffman LH, Fu R, et al. Preventive Services Task Force. Genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility: Systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2005;143: 362–379.
51. Taylor N, Mutch DG. Gynecologic manifestations of hereditary nonpolyposis colorectal cancer. From inherited to sporadic disease. *Oncology (Huntington)*. 2006; 20:85–94; discussion 94–96, 100 review.
52. Offit K, Kauff ND. Reducing the risk of gynecologic cancer in the Lynch syndrome. *N Engl J Med*. 2006;354: 293–295.
53. Schorge JO, et al. SGO White paper on ovarian cancer: etiology, screening and surveillance. *Gynecol Oncol*. 2010;119: 7–17.
54. Boyd J, Rubin SC. Hereditary ovarian cancer: molecular genetics and clinical implications. *Gynecol Oncol*. 1997;71(2, suppl):517–523.
55. O'Donovan PJ, Livingston DM. BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair. *Carcinogenesis*. 2010;31:961.
56. Whittemore AS, Gong G, Itnyre J. Prevalence and contribution of BRCA1 mutations in breast cancer and ovarian cancer: result from three U.S. population-based case-control studies of ovarian cancer. *Am J Hum Genet*. 1997;60: 496–504.
57. Ford D, Easton DP, Stratton M, et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *Am J Hum Genet*. 1998;62:676–689.
58. Holschneider CH, Berek J. Ovarian cancer: epidemiology, biology, and prognostic factors. *Semin Surg Oncol*. 2000; 19:3–10.
59. Antoniou AC, Wang X, Fredericksen ZS, et al. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nat Genet*. 2010;42: 885–892.
60. Pruthi S, Gostout BS, Lindor NM. Identification and management of women with BRCA mutations or hereditary predisposition for breast and ovarian cancer. *Mayo Clin Proc*. 2010;85: 1111–1120.
61. Lu KH, Broaddus RR. Gynecologic cancers in Lynch syndrome/HNPCC. *Fam Cancer*. 2005;4:249–254.

Case 3:16-md-02738-MAS-RLS Document 9732-4 Filed 05/07/19 Page 41 of 200 PageID: 33896

62. Lynch HT, Lynch PM, Lanspa SJ, et al. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin Genet*. 2009;76: 1–18.
63. Lu KH, Dinh M, Kohlmann W, et al. Gynecologic cancer as a ‘‘sentinel cancer’’ for women with hereditary nonpolyposis colorectal cancer syndrome. *Obstet Gynecol*. 2005;105:569–574.
64. Whiteman DC, Siskind V, Purdie DM, et al. Timing of pregnancy and the risk of epithelial ovarian cancer. *Cancer Epidemiol Biomarkers Prev*. 2003;12:42–46.
65. Rostgaard K, Wohlfahrt J, Andersen PK, et al. Does pregnancy induce the shedding of premalignant ovarian cells? *Epidemiology*. 2003;14:168–173.
66. Tworoger SS, Fairfield KM, Colditz GA, et al. Association of oral contraceptive use, other contraceptive methods, and infertility with ovarian cancer risk. *Am J Epidemiol*. 2007;166:894–901.
67. Rossing MA, Tang MT, Flagg EW, et al. A case-control study of ovarian cancer in relation to infertility and the use of ovulation-inducing drugs. *Am J Epidemiol*. 2004;160:1070–1078.
68. Jensen A, Sharif H, Frederiksen K, et al. Use of fertility drugs and risk of ovarian cancer: Danish population based cohort study. *Br Med J*. 2009;338:b249.
69. Mahdavi A, Pejovic T, Nezhat F. Induction of ovulation and ovarian cancer: a critical review of the literature. *Fertil Steril*. 2006;85:819.
70. Weiss NS, Cook LS, Farrow DC, et al. Ovarian cancer. In: Schottenfeld D, Fraumeni JF Jr. *Cancer Epidemiology and Prevention*. New York, NY: Oxford University Press; 1996: 1040–1057.
71. Beral V, Doll R, Hermon C, et al. Collaborative Group on Epidemiological Studies of Ovarian CancerOvarian cancer and oral contraceptives: collaborative reanalysis of data from epidemiological studies including 23,257 women with ovarian cancer and 87,303 controls. *Lancet*. 2008;371:303–314.
72. Narod SA, Risch H, Moslehi R, et al. Oral contraceptives and the risk of hereditary ovarian cancer. Hereditary Ovarian Cancer Clinical Study Group. *N Engl J Med*. 1998;339:424–428.
73. Danforth KN, Tworoger SS, Hecht JL, et al. Breastfeeding and risk of ovarian cancer in two prospective cohorts. *Cancer Causes Control*. 2007;18:517–523.
74. Titus-Ernstoff L, Rees JR, Terry KL, et al. Breast-feeding the last born child and risk of ovarian cancer. *Cancer Causes Control*. 2010;21:201–207.
75. Jordan SJ, Siskind V, Green C. Breastfeeding and risk of epithelial ovarian cancer. *Cancer Causes Control*. 2010;21: 109–116.
76. Hoekstra A, Rodriguez GC. Chemoprevention of ovarian cancer. *Cancer Treat Res*. 2009;149:3–34.
77. Choi JH, et al. Gonadotropins and ovarian cancer. *Endocr Rev*. 2007;28: 440–461.
78. Canman CE, Chen C-Y, Lee MH, et al. *Cold Spring Harbor Symposia on Quantitative Biology, LIX DNA Damage Responses: p53 Induction, Cell Cycle Pertubations, and Apoptosis*. Woodbury, NY: Cold Spring Harbor Press; 1994: 277–286.
79. Ponzoni M, Bocca P, Chiesa V, et al. Differential effects of *N*-(4-hydroxyphenyl) retinamide and retrinoic acid on neuroblastoma cells: apoptosis versus differentiation. *Can Res*. 1995;55: 853–861.
80. Delia D, Aiello A, Lombardi L, et al. *N*-(4-hydroxyphenyl) retinamide induces apoptosis of malignant hemopoietic cell lines including those unresponsive to retinoic acid. *Cancer Res*. 1993;53: 6036–6041.
81. Lotan R. Retinoids in cancer chemoprevention. *FASEB J*. 1996;10: 1031–1039.
82. Kuo SM. Antiproliferative potency of structurally distinct dietary flavonoids on human colon cancer cells. *Cancer Lett*. 1996;110:41–48.
83. Thompson HJ, Jiang C, Lu J, et al. Sulfone metabolite of sulindac inhibits mammary carcinogenesis. *Cancer Res*. 1997;57:420–425.
84. Gould MN. Cancer chemoprevention and therapy by monoterpenes [review] .

Environ Health Perspect. 1997;105 (suppl 4):977–979.
85. Pascale RM, Simile MM, De Miglio MR, et al. Chemoprevention by *S*-adenosyl-L-methionine of rat liver carcinogenesis initiated by 1,2-dimethyl-hydrazine and promoted by orotic acid. *Carcinogenesis*. 1995;16:427–430.
86. el-Bayoumy K, Upadhyaya P, Chae YH, et al. Chemoprevention of cancer by organoselenium compounds [review]. *J Cell Biochem*. 1995;22 (suppl):92–100.
87. Sun SY, Hail N, Lotan R. Apoptosis as a novel target for cancer prevention. *J Natl Cancer Inst*. 2004;96:662–672.
88. Ouyang N, Williams JL, Rigas B. NO-donating aspirin isomers downregulate peroxisome proliferator-activated receptor (PPAR)delta expression in APC (min/+) mice proportionally to their tumor inhibitory effect: implications for the role of PPARdelta in carcinogenesis. *Carcinogenesis*. 2006;27: 232–239.
89. Sinicrope FA, Half E, Morris JS, et al. Cell proliferation and apoptotic indices predict adenoma regression in a placebo-controlled trial of celecoxib in familial adenomatous polyposis patients. *Cancer Epidemiol Biomarkers Prev*. 2004;13:920–927.
90. Vereide AB, Kaino T, Sager G, et al. Bcl-2, BAX, and apoptosis in endometrial hyperplasia after high dose gestagen therapy: a comparison of responses in patients treated with intrauterine levonorgestrel and systemic medroxyprogesterone. *Gynecol Oncol*. 2005;97: 740–750.
91. Reed JC. Mechanisms of apoptosis avoidance in cancer. *Curr Opin Oncol*. 1999;11:68–75.
92. Schildkraut JM, Calingaert B, March banks PA, et al. Impact of progestin and estrogen potency in oral contraceptives on ovarian cancer risk. *J Natl Cancer Inst*. 2002;94:32–38.
93. Lacey JV, Mink PJ, Lubin JH, et al. Menopausal hormone replacement therapy and risk of ovarian cancer. *J Am Med Assoc*. 2002;288:334–341.
94. Rodriguez C, Patel AV, Calle EE, et al. Estrogen replacement therapy and ovarian cancer mortality in a large prospective study of US women. *J Am Med Assoc*. 2001;285:1460–1465.
95. Coughlin SS, Giustozzi A, Smith SJ, et al. A meta-analysis of estrogen replacement therapy and the risk of epithelial ovarian cancer. *J Clin Epidemiol*. 2000;53:367–375.
96. Garg P, Kerlikowdke K, Subak L, et al. Hormone replacement therapy and the risk of epithelial ovarian carcinoma: a meta-analysis. *Obstet Gynecol*. 1998;92: 472–479.
97. Riman T, Dickman PW, Nilsson S, et al. Hormone replacement therapy and the risk of invasive epithelial ovarian cancer in 98. Swedish women. *J Natl Cancer Inst*. 2002;94:497–504.
98. Purdie DM, Bain CJ, Siskind V, et al. Hormone replacement therapy and risk of epithelial ovarian cancer. *Br J Cancer*. 1999;81:559–563.
99. Yang HP, Trabert B, Murphy MA, et al. Ovarian cancer risk factors by histologic subtypes in the NIH-AARP diet and health study. *Int J Cancer*. 2011. doi: 10.1002/ijc.26469.
100. Modugno F. Ovarian cancer and polymorphisms in the androgen and progesterone receptor genes: a HuGE review. *Am J Epidemiol*. 2004;159:319–335.
101. Olsen CM, Green AC, Nagle CM, et al. Epithelial ovarian cancer: testing the "androgens hypothesis". *Endocr Relat Cancer*. 2008;4:1061–1068.
102. Cottreau CM, Ness RB, Modugno F, et al. Endometriosis and its treatment with danazol or lupron in relation to ovarian cancer. *Clin Cancer Res*. 2003;9: 5142–5144.
103. Greer JB, Modugno F, Allen GO, et al. Androgenic progestins in oral contraceptives and the risk of epithelial ovarian cancer. *Obstet Gynecol*. 2005;105: 731–740.
104. Ness RB, Cottreau C. Possible role of ovarian epithelial inflammation in ovarian cancer. *J Natl Cancer Inst*. 1999;91: 1459–1467. Review.
105. Shan W, Liu J. Inflammation: a hidden path to breaking the spell of ovarian cancer. *Cell Cycle*. 2009;8:3107–3111.

106. McSorley MA, Alberg AJ, Allen DS, et al. C-reactive protein concentrations and subsequent ovarian cancer risk. *Obstet Gynecol*. 2007;109:933–941.
107. Toriola AT, Grankvist K, Agborsangaya CB, et al. Changes in pre-diagnostic serum C-reactive protein concentrations and ovarian cancer risk: a longitudinal study. *Ann Oncol*. 2011;22:1916–1921.
108. Clendenen TV, Lundin E, Zeleniuch-Jacquotte A, et al. Circulating inflammation markers and risk of epithelial ovarian cancer. *Cancer Epidemiol Biomarkers Prev*. 2011;20:799–810.
109. Jensen JT, Speroff L. Health benefits of oral contraceptives. *Obstet Gynecol Clin North Am*. 2000;27:705–721.
110. Henderson WJ, Hamilton RC, Griffiths K. Talc in normal and malignant ovarian tissue. *Lancet*. 1979;1:499.
111. Cramer DW, Welch WR, Scully RE, et al. Ovarian cancer and talc: a case-control study. *Cancer*. 1982;50:372–376.
112. Rosenblatt KA, Szklo M, Rosenshein NB. Mineral fiber exposure and the development of ovarian cancer. *Gynecol Oncol*. 1992;45:20–25.
113. Chang S, Risch HA. Perineal talc exposure and risk of ovarian carcinoma. *Cancer*. 1997;79:2396–2401.
114. Harlow BL, Cramer DW, Bell DA, et al. Perineal exposure to talc and ovarian cancer risk. *Obstet Gynecol*. 1992;80:19–26.
115. Huncharek M, Geschwind JF, Kupelnick B. Perineal application of cosmetic talc and risk of invasive epithelial ovarian cancer: a meta-analysis of 11,933 subjects from sixteen observational studies. *Anticancer Res*. 2003;23:1955–1960.
116. Sayasneh A, Tsivos D, Crawford R. Endometriosis and ovarian cancer: a systematic review. *ISRN Obstet Gynecol*. 2011;2011:140310.
117. Ness RB. Endometriosis and ovarian cancer: thoughts on shared pathophysiology. *Am J Obstet Gynecol*. 2003;189:280–294. Review.
118. Al-Sabbagh M, Lam EW, Brosens JJ. Mechanisms of endometrial progesterone resistance. *Mol Cell Endocrinol*. 2011. [Epub ahead of print].
119. Kondi-Pafiti A, Papakonstantinou E, Iavazzo C, et al. Clinicopathological characteristics of ovarian carcinomas associated with endometriosis. *Arch Gynecol Obstet*. 2011. [Epub ahead of print].
120. Risch HA, Howe GR. Pelvic inflammatory disease and the risk of epithelial ovarian cancer. *Cancer Epidemiol Biomarkers Prev*. 1995;4:447–451.
121. Lin HW, Tu YY, Lin SY, et al. Risk of ovarian cancer in women with pelvic inflammatory disease: a population-based study. *Lancet Oncol*. 2011;12:900–904.
122. Cibula D, Widschwendeter M, Majek O, et al. Tubal ligation and the risk of ovarian cancer: review and meta-analysis. *Hum Reprod Update*. 2011;17:55–67.
123. Narod SA, Sun P, Ghadirian P, et al. Tubal ligation and risk of ovarian cancer in carriers of BRCA1 or BRCA2 mutations: a case-control study. *Lancet*. 2001;357:1467–1470.
124. Dennerstein L, Gotts G, Brown JV. Effects of age and non-hormonal contraception on menstrual cycle characteristics. *Gynecol Endocrinol*. 1997;11:127–133.
125. Olsen CM, Green AC, Whiteman DC, et al. Obesity and the risk of epithelial ovarian cancer: a systematic review and meta-analysis. *Eur J Cancer*. 2007;43:690–709.
126. Calle EE, Rodriguez C, Walker-Thurmond K, et al. Overweight, obesity and mortality from cancer in prospectively studied cohort of US adults. *N Engl J Med*. 2003;348:1625–1638.
127. Fairfield KM, Willett WC, Rosner BA, et al. Obesity weight gain, and ovarian cancer. *Obstet Gynecol*. 2002;100:288–296.
128. Purdie DM, Siskind V, Bain CJ, et al. Body size and ovarian cancer: case-control study and systematic review. *Cancer Causes Control*. 2001;12:855–863.
129. Leitzmann MF, Koebnick C, Danforth KN, et al. Body mass index and risk of ovarian cancer. *Cancer*. 2009;115:812–822.
130. McLemore MR, et al. Epidemiologic and genetic factors associated with

ovarian cancer. *Cancer Nursing*. 2009; 32:281–290.
131. LaVecchia C, Decarli A, Franceschis NE, et al. Dietary factors and the risk of epithelial ovarian cancer. *J Natl Cancer Inst*. 1987;79:663–669.
132. Kushi LH, Mink PJ, Sellers TA, et al. Prospective study of diet and ovarian cancer. *Am J Epidemiol*. 1999;149: 21–31.
133. Bosetti C, Negri E, Francschi S, et al. Olive oil, seed oils and other added fats in relation to ovarian cancer (Italy). *Cancer Causes Control*. 2002;13: 465–470.
134. Bosetti C, Negri E, Francschi S, et al. Diet and ovarian cancer risk: a case-control study in Italy. *Int J Cancer*. 2001;93:911–915.
135. Koushik A, Hunter DJ, Spiegelman D, et al. Fruits and vegetables and ovarian cancer risk in a pooled analysis of 12 cohort studies. *Cancer Epidemiol Biomarkers Prev*. 2005;14:2160–2167.
136. Schulz M, Lahmann PH, Boeing H, et al. Fruit and vegetable consumption and risk of epithelial ovarian cancer: The European Prospective Investigation into Cancer and Nutrition. *Cancer Epidemiol*. 2005;14:2531–2535.
137. Larsson SC, Bergkvist L, Wolk A. Milk and lactose intakes and ovarian cancer risk in the Swedish mammography cohort. *Am J Clin Nutr*. 2004;80: 1353–1357.
138. Mommers M, Schouten LJ, Goldbohm RA, et al. Dietary consumption and ovarian cancer risk in the Netherlands Cohort Study on Diet and cancer. *Br J Cancer*. 2006;94:165–170.
139. Chang ET, Lee VS, Canchola AJ, et al. Dietary patterns and risk of ovarian cancer in the California Teachers Study cohort. *Nutr Cancer*. 2008;60:285–291.
140. Chandran U, Bander EV, Olson SH, et al. Healthy eating index and ovarian cancer risk. *Cancer Causes Control*. 2011;22:563–571.
141. Bertone ER, Newcomb PA, Willett WC, et al. Recreational physical activity and ovarian cancer in a population-based case-control study. *Int J Cancer*. 2002; 99:431–436.
142. Tavani A, Gallus S, La Vecchia C, et al. Physical activity and risk of ovarian cancer: an Italian case-control study. *Int J Cancer*. 2001;91:407–411.
143. Bertone ER, Willett WC, Rosner BA, et al. Prospective study of recreational physical activity and ovarian cancer. *J Natl Cancer Inst*. 2001;93:942–948.
144. Anderson JP, Ross JA, Folsom AR. Anthropometric variables, physical activity, and incidence of ovarian cancer: The Iowa Women's Health Study. *Cancer*. 2004;100:1515–1521.
145. Pan SY, Ugnat AM, Mao Y. Physical activity and the risk of ovarian cancer: a case-control study in Canada. *Int J Cancer*. 2005;117:300–307.
146. Biesma RG, Scouten LJ, Dirx MJ, et al. Physical activity and the risk of ovarian cancer: results from the Netherlands Cohort Study Cancer Causes control. 2006;17:109–115.
147. Tworoger SS, Gertig DM, Gates MA, et al. Caffeine, alcohol, smoking, and the risk of incident epithelial ovarian cancer. *Cancer*. 2008;112:1169–1177.
148. Marchbanks PA, Wilson H, Bastos E, et al. Cigarette smoking and epithelial ovarian cancer by histologic type. *Obstet Gynecol*. 2000;95:255–260.
149. Kurian AW, Balise RR, McGuire V, et al. Histologic types of epithelial ovarian cancer: have they different risk factors? *Gynecol Oncol*. 2005;96: 520–530.
150. Jordan SJ, Whiteman DC, Purdie DM, et al. Does smoking increase risk of ovarian cancer? A systematic review. *Gynecol Oncol*. 2006;103:1122–1129.
151. Goodman MT, Tung KH. Active and passive tobacco smoking and the risk of borderline and invasive ovarian cancer. *Cancer Causes Control*. 2003;14: 569–577.
152. Zenzes MT, Reed TE, Casper RF. Effect of cigarette smoking and age on the maturation of human oocytes. *Hum Reprod*. 1997;12:1736–1741.
153. Weshoff C, Gentile F, Lee J, et al. Predictors of ovarian steroid secretion in reproductive age women. *Am J Epidemiol*. 1996;144:381–388.

154. Khaw K, Tazuke S, Barrett-Connor E. Cigarette smoking and levels of adrenal androgens in postmenopausal women. *N Engl J Med*. 1988;318:1705–1709.
155. McKinlay SM. The normal menopause transition: an overview. *Maturitas*. 1996;23:137–145.
156. Garland CF, Mohr SB, Gorham ED, et al. Role of ultraviolet B irradiance and vitamin D in prevention of ovarian cancer. *Am J Prev Med*. 2006;31:512–514.
157. Lefkowitz ES, Garland CF. Sunlight, vitamin D, and ovarian cancer mortality rates in US women. *Int J Epidemiol*. 1994;23:1133–1136.
158. Choi KC, et al. The regulation of apoptosis by activin and transforming growth factor-beta in early neoplastic and tumorigenic ovarian surface epithelium. *J Clin Endocrinol Metab*. 2001;86:2125–2135.
159. Jiang F, Bao J, Li P, et al. Induction of ovarian cancer cell apoptosis by 1,25-dihydroxyvitamin D3 through the down-regulation of telomerase. *J Biol Chem*. 2004;279:53213–53221.
160. Salazar-Martinez E, et al. Nutritional determinants of epithelial ovarian cancer risk: a case-control study in Mexico. *Oncology*. 2002;63:151–157.

Exhibit G

Case 3:16-md-02738-MAS-RLS Document 9732-4 Filed 05/07/19 Page 47 of 200 PageID: 33902

Risk Factors for Ovarian Carcinoma

Adrianne R. Mallen, MD[a], Mary K. Townsend, ScD[b], Shelley S. Tworoger, PhD[b,c,*]

KEYWORDS

• Ovarian cancer • Risk factors • Descriptive epidemiology • Risk reduction • Tumor heterogeneity

KEY POINTS

- Ovarian cancer continues to be the leading gynecologic killer of women in the United States.
- Most women present with advanced-stage disease at time of diagnosis and there are currently no effective screening strategies for average-risk women.
- Cancer epidemiology greatly contributes to the understanding of factors that may modify disease development and drive tumor heterogeneity.

INTRODUCTION

Ovarian cancer is the second most common gynecologic malignancy overall worldwide and the most lethal gynecologic malignancy in the United States and Europe. Each year, approximately 200,000 women worldwide are diagnosed with ovarian cancer and approximately 125,000 women die from the disease.[1] Most patients present with advanced-stage disease because symptoms of early-stage disease may be subtle or generalized.[2] Standard treatment of advanced ovarian cancer involves cytoreductive surgery in combination with taxane-platinum–based chemotherapy.[1] However, most patients experience recurrence and eventually succumb to their disease even with optimal initial treatment.[3]

Given this, identifying risk factors, preventive strategies, and high-risk populations is crucial. However, epidemiologic studies face several challenges. First, ovarian cancer is rare. Furthermore, because ovarian cancer is a heterogeneous disease, considering

Disclosure: The authors report no disclosures.
[a] Department of Gynecologic Oncology, Moffitt Cancer Center/University of South Florida, 12902 Magnolia Drive MCC-GME, Tampa, FL 33612, USA; [b] Department of Cancer Epidemiology, Moffitt Cancer Center, 12902 Magnolia Drive MCC-CANCONT, Tampa, FL 33612, USA; [c] Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA
* Corresponding author. Moffitt Cancer Center, 12902 Magnolia Drive MRC-CANCONT, Tampa, FL 33612.
E-mail address: Shelley.Tworoger@moffitt.org

Hematol Oncol Clin N Am ■ (2018) ■–■
https://doi.org/10.1016/j.hoc.2018.07.002
hemonc.theclinics.com
0889-8588/18/© 2018 Elsevier Inc. All rights reserved.

Downloaded for Anonymous User (n/a) at School of Nursing from ClinicalKey.com by Elsevier on October 02, 2018.
For personal use only. No other uses without permission. Copyright ©2018. Elsevier Inc. All rights reserved.

outcomes of specific cancer subtypes is critical to provide clues to underlying mechanisms. As a result, it is crucial to have large sample sizes to ensure power. Thus, several consortia have been initiated to pool resources from multiple studies and conduct investigations that would not be possible in any single study. Pooling studies that span different time periods further allows addressing a second challenge, which is the temporal changes in clinical characterization of ovarian cancer and changes in certain exposures (eg, oral contraceptive pill [OCP] doses) over time.

Importantly, removal of the ovaries and fallopian tubes reduces risk by up to 80% to 90%.[4] However, negative health consequences, including cardiovascular mortality,[5,6] necessitate the use of this procedure only among high-risk women who would have a net benefit, such as those with *BRCA* or other high-penetrance mutations. However, in average-risk women, efforts to develop well-calibrated risk prediction models have been largely unsuccessful, with low predictive capability even when using known ovarian cancer risk factors (area under the curve [AUC], 0.59–0.64).[7–10] Addition of low-penetrance alleles only modestly improved the AUC to 0.66,[11] requiring identification of new risk factors.[12] A potential reason for the low predictive ability is ovarian cancer heterogeneity, necessitating consideration of subtype-specific risk factor associations. The focus of this article is to review risk factor associations by tumor subtypes to inform the future research that is needed to improve risk prediction.

NONEPITHELIAL OVARIAN CANCER RISK FACTORS

A small proportion of ovarian tumors are from a nonepithelial origin and generally have not been considered in risk modeling efforts. Specifically, sex-cord stromal ovarian neoplasms represent only 1.2% of ovarian cancer cases. These tumors are diagnosed at earlier stages and younger ages, in sharp contrast with epithelial ovarian cancer.[13] Limited data suggest that nonwhite, obese women with a family history of breast or ovarian cancer are at increased risk for this subtype. *BRCA* germline mutations or a genetic predisposition to breast cancer are not related,[14] although germline mutations in *DICER1*[15] and somatic mutations in *FOXL2* are related to these tumors.[16] Ovarian germ cell tumors account for 5% of malignant ovarian neoplasms,[17] with early stage at younger ages.[18] The incidence increases around puberty.[19] There is a greater incidence among Asian/Pacific Islander and Hispanic women than in white women.[20] No definite genetic abnormalities have been identified in families with germ cell tumors.

EPITHELIAL OVARIAN CANCER RISK FACTORS

Epithelial ovarian cancer comprises greater than 90% of malignant epithelial neoplasms and often is diagnosed in postmenopausal women. Incidence is higher in white women (12.8 per 100,000) than in black women (9.8 per 100,000).[21] Incidence seems to be lowest for American Indians/Alaska Natives. Incidence has been declining, with a 1.6% decrease in incidence and 2.1% decrease in mortality per year from 2003 to 2012 in the United States.[22]

Many traditional ovarian cancer risk factors are reproductive or hormonal. In general, processes that decrease the number of ovulatory cycles are protective. For example, OCP use, multiparity, breastfeeding, and tubal ligation, as well as late age at menarche and early age at menopause, have been consistently associated with decreased risk, many with a dose-response relationship.[22] However, studies among women using more recent lower-dose OCP formulations do not observe a decreased risk except with very long durations of use (>10 years).[23–25] Further, use of hormone therapy, including unopposed estrogen and combined estrogen and progestin, seems

Downloaded for Anonymous User (n/a) at School of Nursing from ClinicalKey.com by Elsevier on October 02, 2018.
For personal use only. No other uses without permission. Copyright ©2018. Elsevier Inc. All rights reserved.

to increase risk.[26–31] Other risk factors include endometriosis, taller height, and high body mass index in adolescence.[32–36]

Variation in Risk Associations according to Ovarian Cancer Subtypes

Ovarian cancers represent a diverse group of diseases that are unique based on precursor lesions, histology, cause, developmental origins, as well as distinct mutational profiles.[37,38] Stratification based on subtypes is critical for understanding mechanisms underlying risk factor associations and for developing improved risk prediction models. Although the most common assessment of heterogeneity is based on histologic subtypes (ie, the morphologic features of the tumor) and grade, other metrics have also been used. Large-scale studies that examined risk factors for specific ovarian cancer subtypes are summarized later.

Histologic subtypes

Unexpectedly, most known ovarian cancer risk factors show stronger associations with nonserous tumors, which comprise ∼25% of epithelial ovarian cancers, than the more aggressive serous tumors (**Table 1**). For example, in a pooled analysis of 21 prospective cohort studies in the Ovarian Cancer Cohort Consortium (OC3), reproductive risk factors, including lower parity and older age at menopause, as well as endometriosis, were associated primarily with increased risks of endometrioid and clear cell tumors.[31] This finding is consistent with pooled analyses of case-control studies and studies of endogenous hormones.[39,40] Notably, OCP use seems equally protective across histologic subtypes in multiple studies.[31,39] Surgical procedures, including tubal ligation and hysterectomy, also seem to primarily decrease the risk of nonserous tumors.[31,41–44] Data on histologic subtype–specific associations for salpingectomy are currently unavailable, because few studies have examined this association and most have had few exposed cases.[31,42,43]

Associations of several lifestyle factors and use of over-the-counter medications with risk of specific ovarian cancer histologic subtypes have also been investigated. Smoking was associated with an increased risk of mucinous ovarian tumors but a decreased risk of clear cell tumors in several studies.[31,45] A pooled analysis of 8 case-control studies found modest increases in risks of serous, endometrioid, and clear cell carcinomas, but not mucinous tumors, in women who used genital talc powder.[46] Aspirin and other nonsteroidal antiinflammatory drug use was mainly associated with serous disease in both prospective and retrospective consortial analyses.[47] Similarly, history of ovarian cancer is one of the few factors that is more strongly associated with serous carcinoma.[31] Family history of breast cancer was most strongly related to endometrioid tumors.

Multiple studies have integrated grade and histologic subtype to evaluate associations for high-grade and low-grade serous tumors separately because these are thought to have different causes.[31,42,43] In general, low-grade serous tumors had similar associations to endometrioid and clear cell disease, although family history of ovarian cancer was related to high-grade serous tumors.[31] A key caveat in these studies is that grade does not have standard classification criteria and is often missing in epidemiologic studies, reducing power and leading to misclassification of disease subtype.

Biologically, these results support the theories of differing cells of origin in ovarian cancer, notably with endometriosis and tubal ligation being strongly associated with histologic subtypes thought to be directly linked with endometriotic tissue and retrograde menstruation.[48] Similarly, the family history of ovarian cancer relationship with high-grade serous disease is likely explained in part via BRCA mutations. In the

Downloaded for Anonymous User (n/a) at School of Nursing from ClinicalKey.com by Elsevier on October 02, 2018.
For personal use only. No other uses without permission. Copyright ©2018. Elsevier Inc. All rights reserved.

Table 1
Summary of putative cells of origin and identified risk factors for specific ovarian cancer histologic subtypes

Subtype	Putative Cells of Origin	Reproductive and Hormonal Risk Factors	Family History, Demographic, and Lifestyle Risk Factors
All serous	Ovarian surface epithelium, fallopian tube epithelium	Lower parity[31,39] Shorter duration of OC use[31,39] HT use/longer duration of use[31,39] No history of tubal ligation[42–44]	Family history of breast cancer[31] Family history of ovarian cancer[31] Taller height[31] Genital powder use[46] No regular aspirin use[47]
High-grade serous	Ovarian surface epithelium, fallopian tube epithelium	Lower parity[31] Shorter duration of OC use[31] Longer duration of HT use[31] No history of tubal ligation[42,43]	Family history of ovarian cancer[31] Taller height[31]
Low-grade serous	Ovarian surface epithelium, fallopian tube epithelium	Lower parity[31] Shorter duration of OC use[31] Longer duration of HT use[31]	—
Endometrioid	Endometriosis	[a]Lower parity[31,39] Shorter duration of OC use[31,39] HT use/longer duration of use[31,39] [a]Older age at menopause[31,39] [a]No history of tubal ligation[31,42–44] Endometriosis[31]	[a]Family history of breast cancer[31] Taller height[31] Genital powder use[46]
Clear cell	Endometriosis	[a]Lower parity[31,39] Shorter duration of OC use[31,39] Shorter duration of HT use[31] [a]Older age at menopause[31,39] [a]No history of tubal ligation[31,42,43] No history of hysterectomy[31] Endometriosis[31]	Taller height[31] Never smoking[31] Genital powder use[46]
Mucinous	Unknown	Lower parity[31,39] No history of tubal ligation[42]	Taller height[31] More pack-years[31,45]

Abbreviations: HT, postmenopausal hormone therapy; OC, oral contraceptive.
[a] Indicates that the risk factor was most strongly related to this subtype(s).

OC3 analysis, unstructured hierarchical clustering suggested that few known risk factors were associated with serous tumors compared with endometrioid and clear cell diseases, which had very similar risk factor profiles.[31] This finding is in stark contrast with breast cancer, for which risk factors for the most common type of tumor (estrogen receptor positive) are well understood, and may explain the poor predictive ability of prior risk models. Focusing on the risk factors that have been identified for serous disease may open up new areas of research to identify novel risk factors to best identify high-risk women and elucidate novel risk-reduction strategies.[49]

Type 1 versus type 2

An additional method of classifying ovarian cancer subtypes groups certain histologic subtypes together based on putative cells of origin and somatic mutations and has been used in risk factor studies to enhance power.[50] Type 1 cancers consist of low-grade serous, endometrioid, clear cell, and mucinous cancers arising from the ovarian

Downloaded for Anonymous User (n/a) at School of Nursing from ClinicalKey.com by Elsevier on October 02, 2018.
For personal use only. No other uses without permission. Copyright ©2018. Elsevier Inc. All rights reserved.

epithelium or endometriosis and are characterized by mutations in *KRAS*, *ARID1A*, *PIK3CA*, *PTEN*, and *BRAF*. Type 2 cancers, which comprise high-grade serous cancers, carcinosarcomas, and undifferentiated carcinomas, are characterized by *TP53* mutations and likely originate from the distal end of the fallopian tube. In general, these studies have observed similar associations to those described earlier when looking at the finer granularity of histologic subtype and grade. For example, reproductive factors such as parity and tubal ligation were most strongly associated with a lower risk of type 1 tumors, whereas OCP use was consistently associated with a lower risk across both types.[39,51,52]

Anatomic site

Research on ovarian cancer has historically encompassed primary ovarian, primary peritoneal, and primary fallopian tube cancers. However, several studies have explored whether risk factor profiles differ by the anatomic site of the cancer, which might imply different carcinogenic origins. Among these studies, most have used case-case designs in which peritoneal or fallopian tube cancer cases were compared with ovarian cancer cases,[53–57] although several studies compared 2 or more case groups defined by site of origin with a common healthy control group,[58,59] allowing direct comparison of odds ratios (ORs) across anatomic sites. Although results are not entirely clear, these studies suggest that associations of several established risk factors may vary by tumor site of origin such that associations with ovarian cancer are in the expected direction, whereas associations with fallopian tube and peritoneal cancers may be similar, null, or in the opposite direction.

For example, in the Australian Ovarian Cancer Study (AOCS), which included invasive serous ovarian (n = 627), peritoneal (n = 129), and fallopian tube cancer cases (N = 45) and 1508 control women, higher parity and longer duration of breastfeeding were each associated with lower risks of ovarian cancer; the associations with fallopian tube cancer were similar to those for ovarian cancer, whereas the associations with peritoneal cancer were null or attenuated.[59] In the North Carolina Ovarian Cancer Study (NCOCS), which enrolled 495 women with epithelial ovarian cancer, 62 women with primary peritoneal cancer, and 1086 control women, ORs for ever being pregnant and number of pregnancies were similarly inverse for ovarian and peritoneal cancers; however, older age at last pregnancy was associated with a decreased risk of ovarian cancer (OR, 0.58; 95% confidence interval [CI], 0.39–0.86 comparing age $\geq$ 35 years vs <25 years), but an increased risk of peritoneal cancer (OR, 2.78; 95% CI, 1.00–7.78). Similarly, tubal ligation was associated with reduced risk of ovarian cancer but not associated with peritoneal cancer in NCOCS, although the RRs were not statistically significantly different. In AOCS, the reduction in risk caused by tubal ligation was similar across anatomic sites.[58]

Given the limited the number of studies, it is difficult to conclude whether cancers at different anatomic sites should be considered distinct outcomes. Continued collaborative efforts are warranted in order to achieve an adequate sample size for continued investigation.

Tumor dominance and laterality

It is now accepted that a substantial proportion of serous tumors arise from the fallopian tubes, whereas some nonserous histologic subtypes, such as endometrioid, may arise from endometriosis or retrograde menstruation. Because ovarian cancer is usually diagnosed at a late stage when disease has spread, determining the cell of origin is often very difficult.[49] Pathology studies have suggested that dominant tumors (restricted to 1 ovary or at least twice as large on 1 ovary compared with the

Downloaded for Anonymous User (n/a) at School of Nursing from ClinicalKey.com by Elsevier on October 02, 2018.
For personal use only. No other uses without permission. Copyright ©2018. Elsevier Inc. All rights reserved.

other) are less likely to have a serous tubal intraepithelial carcinoma and are more likely to be of nonserous histologic subtypes, compared with those with tumor spread more evenly or diffusely across the peritoneal cavity. Further, endometriosis is often found on the left side; this may reflect greater ovulation events on the right side, leading to higher localized progesterone production, which suppresses endometriosis, as well as less efficient elimination of retrograde menstruation caused by anatomic proximity with the colon or decreased flow of peritoneal fluid on the left.[34] Thus, laterality of dominant tumors may be more likely to be related to this cell of origin.

Specifically, in a study of 1386 tumors, nondominant tumors were more likely to be serous and stage III/IV. In addition, nondominant tumors were associated with BRCA 1/2 mutation carrier status, higher parity, and use of estrogen hormone therapy. The association with BRCA mutations supports the now accepted theory that the distal fallopian tube is the site of high-grade serous cancers among BRCA mutation carriers.[60] In another study among 1771 patients with invasive epithelial ovarian cancer, 61% were dominant, whereas 39% were nondominant. Reproductive factors such as tubal ligation, 2 or more births, endometriosis, and age were more strongly associated with dominant tumors than nondominant tumors,[61] again supporting the role of reproductive factors in tumors with a non–fallopian tube site of origin. These large studies provide provocative evidence of different developmental pathways of ovarian tumors based on a woman's risk factor profile.[60,61]

Tumor aggressiveness

There is wide variation in length of ovarian cancer survivorship. Surveillance, Epidemiology, and End Results (SEER) data from 1998 to 2007 indicated that 47.1% of patients died of ovarian cancer within 3 years of diagnosis versus 34.1% of patients who survived longer than 10 years after diagnosis. In a combined analysis of 4 studies (2 cohort and 2 case control) with a total of 4342 ovarian cases, cases were classified as being rapidly fatal (ie, death within 3 years) or less aggressive disease (all others). Older age (positive association) and OCP use (protective association) were more strongly associated with rapidly fatal than less aggressive disease. Higher parity was only associated with a decreased risk of less aggressive disease. Results were consistent after accounting for differences in study design, geographic location, and timing across cohorts, although sparse data on tumor grade and treatment prevented rigorous consideration of these factors in analyses. Overall, these results may contribute to development of primary prevention strategies for the most aggressive cancers.[35]

GENETIC MUTATIONS AND PREDISPOSITION

Family history remains one of the strongest risk factors for epithelial ovarian cancer. Women with a first-degree relative with ovarian cancer have a 3-fold increased risk of developing the disease compared with women with no family history. Twin studies indicate that inherited genetics are more significant than environmental and lifestyle factors.[62] *BRCA1* and *BRCA2* gene mutations are high-penetrant susceptibility genes and the most influential predictors of inherited risk for ovarian cancer. About 15% of patients with high-grade serous epithelial ovarian cancer have a germline mutation in one of the *BRCA* genes.[63] Women with BRCA mutations almost exclusively develop serous histologic subtype disease.[41] Consistent with this pattern, family histories of breast and ovarian cancer were each associated with an increased risk of serous tumors in the OC3. Family history of breast cancer was also associated with endometrioid carcinomas.[31] The overall risk of ovarian cancer for a woman with a *BRCA1*

Downloaded for Anonymous User (n/a) at School of Nursing from ClinicalKey.com by Elsevier on October 02, 2018.
For personal use only. No other uses without permission. Copyright ©2018. Elsevier Inc. All rights reserved.

mutation is approximately 39% to 46% and 10% to 27% for *BRCA2* mutation carriers by age 70 years.[64–67] In the general population, the estimated risk of carrying a *BRCA* mutation varies between 1 in 300 and 1 in 800 individuals. However, in certain populations, such as Ashkenazi Jews, the mutations are found more frequently in about 1 in 40 individuals. Risk-reducing surgery for known *BRCA* carriers by bilateral salpingo-oophorectomy has been successful in reducing epithelial ovarian cancer mortality. Typically, surgery is recommended for *BRCA1* carriers aged 35 to 40 years and *BRCA2* carriers aged 40 to 45 years, taking into account the patient's future childbearing preferences.[41]

More recent evidence indicates that methylation of the BRCA1 promoter in white blood cells (WBCs) is an additional factor influencing ovarian cancer risk. An analysis of blood samples obtained from 1541 women with ovarian cancer before chemotherapy and 3682 matched controls found that most of the women, regardless of case-control status, had normal germline *BRCA1* test results. However, 9% of women with cancer had abnormal methylation in the *BRCA1* promoter in circulating WBCs compared with 4% of control participants. After adjusting for multiple factors, the presence of methylated *BRCA1* conferred a 3-fold higher risk of ovarian cancer. If confirmed in prospective studies, systemic abnormal promoter methylation of BRCA could be one of the strongest known risk factors beyond germline BRCA mutations.[68] Further, understanding of its relationship to different histologic subtypes of disease would also elucidate the cause of ovarian carcinogenesis.

All the known susceptibility alleles that have currently been identified account for less than half of the heritable component of ovarian cancer, suggesting there are more mutations to be discovered. Although clinical management of *BRCA* mutation carriers is clear, clinical difficulties arise when counseling patients with intermediate-risk susceptibility genes. These genes include *FANCM*, *RAD51C*, *RAD51D*, *BRIP1*, and DNA mismatch repair genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*). The DNA mismatch repair genes are associated with the autosomal dominant, inherited Lynch syndrome, which confers greater risk of gynecologic cancers, with endometrial cancer remaining the most common, but also an increased risk of ovarian cancer. Women with Lynch syndrome who develop ovarian cancer typically have nonserous histology with endometrioid and clear cell tumors as the most common subtypes. Epithelial ovarian cancer risk is estimated to be 4% to 20% in *MLH1* carriers, 7.5% to 20% in *MSH2* carriers, and up to 13.5% in *MSH6* carriers. *PMS2* mutations account for very few cases. Genome-wide association studies have identified 39 independent epithelial ovarian cancer risk regions, with each risk region associated with only modest increased risk. All of these alleles have been associated with high-grade serous epithelial ovarian cancer. In contrast with high-penetrant genes, most of these common variant risk alleles are located in the non–protein-coding regions of the genome, implying that epigenomic regulation of 1 or more target genes is necessary and that they are not directly involved in DNA repair.[63] However, OncoArray and the Collaborative Oncological Gene-Environment Study (OCAC) identified 30 epithelial ovarian cancer risk loci by genome-wide association studies and examined their associations with specific histologic subtypes. They found that *HOXD9* is a likely target susceptibility gene in both serous and mucinous histologic subtypes that also affects focal adhesion within a cancer-related pathway. *HNF1B* was downregulated in most serous ovarian cancers, but overexpressed in clear cell ovarian carcinomas.[69] Histologic subtype–specific studies such as this one will help further the understanding of risk reduction given the heterogeneity of ovarian cancer.

Downloaded for Anonymous User (n/a) at School of Nursing from ClinicalKey.com by Elsevier on October 02, 2018.
For personal use only. No other uses without permission. Copyright ©2018. Elsevier Inc. All rights reserved.

ARTICLE IN PRESS

SUMMARY AND RECOMMENDATIONS

This article indicates that, although epidemiologic studies have made strides in elucidating variations in risk factor profiles according to several classifications of ovarian cancer subtypes, much work is yet to be done to yield results that will shift clinical practice. Current risk prediction models are not accurate enough to factor into decisions about preventive treatment strategies. Following are several recommended research priorities for epidemiologic studies to move closer toward clinical translation potential.

Studies focused on understanding the genetic architecture of ovarian cancer, and particularly ovarian cancer subtypes, are critical to establish effective risk-reduction models. Further, research that goes beyond germline mutations to consider methylation and other DNA modifications, as well as downstream phenomena such as RNA transcription, proteomics, and metabolomics, may be a fruitful approach to better characterizing the variable role of genetics in ovarian carcinogenesis.

In addition, to complement gains in knowledge about the genetics of ovarian cancer, an important focus of epidemiologic research is discovery of novel nongenetic risk factors, especially with regard to high-grade serous ovarian carcinoma, the most common subtype with the most aggressive behavior but the least understood risk factor profile. A more comprehensive understanding of the underlying biology linking risk factors with specific disease subtypes will be critical for developing targeted preventive interventions for women at high risk of ovarian cancer. This work has already begun, with research examining psychosocial factors, environmental exposures, and inflammation, among other factors. For example, there is evidence that C-reactive protein may be more strongly related to risk of serous than nonserous cancer.[70] However, to better elucidate these subtype-specific associations, larger consortial studies are needed and thus greater collaboration among investigators and institutions.

Further, investigators should consider whether the tumor subtype classifications discussed in this article are optimal for clustering subtypes with a common cause, or whether different approaches are warranted. It is possible that traditional disease classification using pathology, molecular characteristics, and survival metrics do not correlate well with tumor developmental biology or the risk factor profiles underlying tumor development. New research focused on investigating the multitude of tumor characteristics (eg, immune markers, microenvironment) will likely uncover new causal factors.

In addition, the ultimate goal of the research recommended here is to improve the ability to prevent ovarian cancer in individual women. Thus, epidemiologists will need to collaborate with scientists in other fields (eg, biostatisticians, data scientists, clinicians) to integrate data on genetics, other omics, and nongenetic risk factors to improve individual-level risk prediction models and identification of women who will benefit most from screening and risk-reducing surgeries.

REFERENCES

1. Shih KK, Chi DS. Maximal cytoreductive effort in epithelial ovarian cancer surgery. J Gynecol Oncol 2010;21:75–80.
2. Vernooij F, Heintz P, Witteveen E, et al. The outcomes of ovarian cancer treatment are better when provided by gynecologic oncologists and in specialized hospitals: a systematic review. Gynecol Oncol 2007;105:801–12.
3. Bristow RE, Puri I, Chi DS. Cytoreductive surgery for recurrent ovarian cancer: a meta-analysis. Gynecol Oncol 2009;112:265–74.

Downloaded for Anonymous User (n/a) at School of Nursing from ClinicalKey.com by Elsevier on October 02, 2018.
For personal use only. No other uses without permission. Copyright ©2018. Elsevier Inc. All rights reserved.

4. Rutter JL, Smith AM, Dávila MR, et al. Mutational analysis of the BRCA1-interacting genes ZNF350/ZBRK1 and BRIP1/BACH1 among BRCA1 and BRCA2-negative probands from breast-ovarian cancer families and among early-onset breast cancer cases and reference individuals. Hum Mutat 2003; 22:121–8.
5. Finch AP, Lubinski J, Møller P, et al. Impact of oophorectomy on cancer incidence and mortality in women with a BRCA1 or BRCA2 mutation. J Clin Oncol 2014;32: 1547–53.
6. Jacoby VL, Grady D, Wactawski-Wende J, et al. Oophorectomy vs ovarian conservation with hysterectomy: cardiovascular disease, hip fracture, and cancer in the Women's Health Initiative Observational Study. Arch Intern Med 2011;171: 760–8.
7. Hartge P, Whittemore AS, Itnyre J, et al. Rates and risks of ovarian cancer in subgroups of white women in the United States. The Collaborative Ovarian Cancer Group. Obstet Gynecol 1994;84:760–4.
8. Li K, Hüsing A, Fortner RT, et al. An epidemiologic risk prediction model for ovarian cancer in Europe: the EPIC study. Br J Cancer 2015;112:1257–65.
9. Pfeiffer RM, Park Y, Kreimer AR, et al. Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: derivation and validation from population-based cohort studies. PLoS Med 2013;10:e1001492.
10. Rosner BA, Colditz GA, Webb PM, et al. Mathematical models of ovarian cancer incidence. Epidemiology 2005;16:508–15.
11. Clyde MA, Palmieri Weber R, Iversen ES, et al. Risk prediction for epithelial ovarian cancer in 11 United States-based case-control studies: incorporation of epidemiologic risk factors and 17 confirmed genetic loci. Am J Epidemiol 2016;184:579–89.
12. Alvarez RD, Karlan BY, Strauss JF. "Ovarian cancers: Evolving paradigms in research and care": report from the Institute of Medicine. Gynecol Oncol 2016; 141:413–5.
13. Quirk JT, Natarajan N. Ovarian cancer incidence in the United States, 1992-1999. Gynecol Oncol 2005;97:519–23.
14. Boyce EA, Costaggini I, Vitonis A, et al. The epidemiology of ovarian granulosa cell tumors: a case-control study. Gynecol Oncol 2009;115:221–5.
15. Heravi-Moussavi A, Anglesio MS, Cheng SW, et al. Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers. N Engl J Med 2012;366:234–42.
16. Nolan A, Joseph NM, Sangoi AR, et al. FOXL2 mutation status in granulosa theca cell tumors of the ovary. Int J Gynecol Pathol 2017;36:568–74.
17. Tewari K, Cappuccini F, Disaia PJ, et al. Malignant germ cell tumors of the ovary. Obstet Gynecol 2000;95:128–33.
18. Zalel Y, Piura B, Elchalal U, et al. Diagnosis and management of malignant germ cell ovarian tumors in young females. Int J Gynaecol Obstet 1996;55:1–10.
19. Moller H, Evans H. Epidemiology of gonadal germ cell cancer in males and females. APMIS 2003;111:43–6 [discussion: 46–8].
20. Smith HO, Berwick M, Verschraegen CF, et al. Incidence and survival rates for female malignant germ cell tumors. Obstet Gynecol 2006;107:1075–85.
21. Goodman MT, Howe HL, Tung KH, et al. Incidence of ovarian cancer by race and ethnicity in the United States, 1992-1997. Cancer 2003;97:2676–85.
22. Tworoger SS, Shafrir AL, Hankinson SE. Ovarian cancer [Chapter 46]. In: Thun M, Linet MS, Cerhan JR, et al, editors. Cancer epidemiology and prevention. 4th edition. New York: Oxford University Press; 2018. p. 889–907.

Downloaded for Anonymous User (n/a) at School of Nursing from ClinicalKey.com by Elsevier on October 02, 2018.
For personal use only. No other uses without permission. Copyright ©2018. Elsevier Inc. All rights reserved.

23. Bethea TN, Palmer JR, Adams-Campbell LL, et al. A prospective study of reproductive factors and exogenous hormone use in relation to ovarian cancer risk among black women. Cancer Causes Control 2017;28:385–91.
24. Koushik A, Grundy A, Abrahamowicz M, et al. Hormonal and reproductive factors and the risk of ovarian cancer. Cancer Causes Control 2017;28:393–403.
25. Shafrir AL, Schock H, Poole EM, et al. A prospective cohort study of oral contraceptive use and ovarian cancer among women in the United States born from 1947 to 1964. Cancer Causes Control 2017;28:371–83.
26. Beral V, Million Women Study, Collaborators, Bull D, et al. Ovarian cancer and hormone replacement therapy in the Million Women Study. Lancet 2007;369: 1703–10.
27. Danforth KN, Tworoger SS, Hecht JL, et al. A prospective study of postmenopausal hormone use and ovarian cancer risk. Br J Cancer 2007;96:151–6.
28. Pearce CL, Chung K, Pike MC, et al. Increased ovarian cancer risk associated with menopausal estrogen therapy is reduced by adding a progestin. Cancer 2009;115:531–9.
29. Reid BM, Permuth JB, Sellers TA. Epidemiology of ovarian cancer: a review. Cancer Biol Med 2017;14:9–32.
30. Trabert B, Wentzensen N, Yang HP, et al. Ovarian cancer and menopausal hormone therapy in the NIH-AARP diet and health study. Br J Cancer 2012;107: 1181–7.
31. Wentzensen N, Poole EM, Trabert B, et al. Ovarian cancer risk factors by histologic subtype: an analysis from the ovarian cancer cohort consortium. J Clin Oncol 2016;34:2888–98.
32. Aune D, Navarro Rosenblatt DA, Chan DS, et al. Anthropometric factors and ovarian cancer risk: a systematic review and nonlinear dose-response meta-analysis of prospective studies. Int J Cancer 2015;136:1888–98.
33. Engeland A, Tretli S, Bjorge T. Height, body mass index, and ovarian cancer: a follow-up of 1.1 million Norwegian women. J Natl Cancer Inst 2003;95:1244–8.
34. Laughlin-Tommaso SK, Stewart EA, Grossardt BR, et al. Incidence, time trends, laterality, indications, and pathological findings of unilateral oophorectomy before menopause. Menopause 2014;21:442–9.
35. Poole EM, Merritt MA, Jordan SJ, et al. Hormonal and reproductive risk factors for epithelial ovarian cancer by tumor aggressiveness. Cancer Epidemiol Biomarkers Prev 2013;22:429–37.
36. Schouten LJ, Goldbohm RA, van den Brandt PA. Height, weight, weight change, and ovarian cancer risk in the Netherlands cohort study on diet and cancer. Am J Epidemiol 2003;157:424–33.
37. Epidemiology Working Group Steering Committee, Ovarian Cancer Association Consortium Members of the EWG SC, in alphabetical order, Doherty JA, Jensen A, Kelemen LE, et al. Current gaps in ovarian cancer epidemiology: the need for new population-based research. J Natl Cancer Inst 2017;109. https://doi.org/10.1093/jnci/djx144.
38. Rosen DG, Yang G, Liu G, et al. Ovarian cancer: pathology, biology, and disease models. Front Biosci (Landmark Ed) 2009;14:2089–102.
39. Fortner RT, Ose J, Merritt MA, et al. Reproductive and hormone-related risk factors for epithelial ovarian cancer by histologic pathways, invasiveness and histologic subtypes: Results from the EPIC cohort. Int J Cancer 2015;137:1196–208.
40. Ose J, Schock H, Poole EM, et al. Pre-diagnosis insulin-like growth factor-I and risk of epithelial invasive ovarian cancer by histological subtypes: a collaborative

Downloaded for Anonymous User (n/a) at School of Nursing from ClinicalKey.com by Elsevier on October 02, 2018.
For personal use only. No other uses without permission. Copyright ©2018. Elsevier Inc. All rights reserved.

re-analysis from the Ovarian Cancer Cohort Consortium. Cancer Causes Control 2017;28:429–35.

41. American College of Obstetricians and Gynecologists, ACOG Committee on Practice Bulletins–Gynecology, ACOG Committee on Genetics & Oncologists, Society of Gynecologic Oncologists. ACOG practice bulletin no. 103: hereditary breast and ovarian cancer syndrome. Obstet Gynecol 2009;113:957–66.
42. Sieh W, Salvador S, McGuire V, et al. Tubal ligation and risk of ovarian cancer subtypes: a pooled analysis of case-control studies. Int J Epidemiol 2013;42: 579–89.
43. Gaitskell K, Green J, Pirie K, et al. Tubal ligation and ovarian cancer risk in a large cohort: Substantial variation by histological type. Int J Cancer 2016;138:1076–84.
44. Cibula D, Widschwendter M, Majek O, et al. Tubal ligation and the risk of ovarian cancer: review and meta-analysis. Hum Reprod Update 2011;17:55–67.
45. Licaj I, Jacobsen BK, Selmer RM, et al. Smoking and risk of ovarian cancer by histological subtypes: an analysis among 300 000 Norwegian women. Br J Cancer 2017;116:270–6.
46. Terry KL, Karageorgi S, Shvetsov YB, et al. Genital powder use and risk of ovarian cancer: a pooled analysis of 8,525 cases and 9,859 controls. Cancer Prev Res (Phila) 2013;6:811–21.
47. Trabert B, Ness RB, Lo-Ciganic WH, et al. Aspirin, nonaspirin nonsteroidal anti-inflammatory drug, and acetaminophen use and risk of invasive epithelial ovarian cancer: a pooled analysis in the Ovarian Cancer Association Consortium. J Natl Cancer Inst 2014;106:djt431.
48. National Academies of Sciences, Engineering, and Medicine. The biology of ovarian cancers. Ovarian cancers: evolving paradigms in research and care [Chapter 2]. Washington, DC: The National Academies Press; 2016.
49. Kurman RJ, Shih Ie M. The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. Am J Surg Pathol 2010;34:433–43.
50. National Academies of Sciences, Engineering, and Medicine. Prevention and early detection. Ovarian cancers: evolving paradigms in research and care [Chapter 3]. Washington, DC: The National Academies Press; 2016.
51. Merritt MA, De Pari M, Vitonis AF, et al. Reproductive characteristics in relation to ovarian cancer risk by histologic pathways. Hum Reprod 2013;28:1406–17.
52. Ose J, Fortner RT, Rinaldi S, et al. Endogenous androgens and risk of epithelial invasive ovarian cancer by tumor characteristics in the European Prospective Investigation into Cancer and Nutrition. Int J Cancer 2015;136:399–410.
53. Barda G, Menczer J, Chetrit A, et al. Comparison between primary peritoneal and epithelial ovarian carcinoma: a population-based study. Am J Obstet Gynecol 2004;190:1039–45.
54. Eltabbakh GH, Piver MS, Natarajan N, et al. Epidemiologic differences between women with extraovarian primary peritoneal carcinoma and women with epithelial ovarian cancer. Obstet Gynecol 1998;91:254–9.
55. Halperin R, Zehavi S, Langer R, et al. Primary peritoneal serous papillary carcinoma: a new epidemiologic trend? A matched-case comparison with ovarian serous papillary cancer. Int J Gynecol Cancer 2001;11:403–8.
56. Schnack TH, Sorensen RD, Nedergaard L, et al. Demographic clinical and prognostic characteristics of primary ovarian, peritoneal and tubal adenocarcinomas of serous histology–a prospective comparative study. Gynecol Oncol 2014;135: 278–84.

Downloaded for Anonymous User (n/a) at School of Nursing from ClinicalKey.com by Elsevier on October 02, 2018.
For personal use only. No other uses without permission. Copyright ©2018. Elsevier Inc. All rights reserved.

57. Sorensen RD, Schnack TH, Karlsen MA, et al. Serous ovarian, fallopian tube and primary peritoneal cancers: a common disease or separate entities - a systematic review. Gynecol Oncol 2015;136:571–81.
58. Grant DJ, Moorman PG, Akushevich L, et al. Primary peritoneal and ovarian cancers: an epidemiological comparative analysis. Cancer Causes Control 2010;21: 991–8.
59. Jordan SJ, Green AC, Whiteman DC, et al. Serous ovarian, fallopian tube and primary peritoneal cancers: a comparative epidemiological analysis. Int J Cancer 2008;122:1598–603.
60. Ivanova A, Loo A, Tworoger S, et al. Ovarian cancer survival by tumor dominance, a surrogate for site of origin. Cancer Causes Control 2015;26:601–8.
61. Kotsopoulos J, Terry KL, Poole EM, et al. Ovarian cancer risk factors by tumor dominance, a surrogate for cell of origin. Int J Cancer 2013;133:730–9.
62. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78–85.
63. Jones MR, Kamara D, Karlan BY, et al. Genetic epidemiology of ovarian cancer and prospects for polygenic risk prediction. Gynecol Oncol 2017;147:705–13.
64. Antoniou A, Pharoah PD, Narod S, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet 2003;72:1117–30.
65. Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol 2007;25:1329–33.
66. Ford D, Easton DF, Stratton M, et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. Am J Hum Genet 1998;62:676–89.
67. King MC, Marks JH, Mandell JB, New York Breast Cancer Study, Group. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science 2003;302:643–6.
68. Lonning PE, Berge EO, Bjørnslett M, et al. White blood Cell BRCA1 promoter methylation status and ovarian cancer risk. Ann Intern Med 2018;168(5):326–34.
69. Kar SP, Berchuck A, Gayther SA, et al. Common genetic variation and susceptibility to ovarian cancer: current insights and future directions. Cancer Epidemiol Biomarkers Prev 2018;27:395–404.
70. Trabert B, Pinto L, Hartge P, et al. Pre-diagnostic serum levels of inflammation markers and risk of ovarian cancer in the prostate, lung, colorectal and ovarian cancer (PLCO) screening trial. Gynecol Oncol 2014;135:297–304.

Downloaded for Anonymous User (n/a) at School of Nursing from ClinicalKey.com by Elsevier on October 02, 2018.
For personal use only. No other uses without permission. Copyright ©2018. Elsevier Inc. All rights reserved.

Exhibit H

Cancer Causes & Control
https://doi.org/10.1007/s10552-018-1082-4

ORIGINAL PAPER

Benign gynecologic conditions are associated with ovarian cancer risk in African-American women: a case–control study

Hyo K. Park[1] · Joellen M. Schildkraut[2] · Anthony J. Alberg[3] · Elisa V. Bandera[4] · Jill S. Barnholtz-Sloan[5] · Melissa Bondy[6] · Sydnee Crankshaw[7] · Ellen Funkhouser[8] · Patricia G. Moorman[7] · Edward S. Peters[9] · Paul Terry[10] · Frances Wang[7] · Julie J. Ruterbusch[1] · Ann G. Schwartz[1] · Michele L. Cote[1]

Received: 26 September 2017 / Accepted: 17 September 2018
© Springer Nature Switzerland AG 2018

Abstract

Background The association between common benign gynecologic conditions and ovarian cancer remains under-studied in African Americans. Therefore, we examine the association between self-reported history of benign gynecologic conditions and epithelial ovarian cancer risk in African-American women.

Methods Data from a large population-based, multi-center case–control study of epithelial ovarian cancer in African-American women were analyzed to estimate the association between self-reported history of endometriosis, pelvic inffammatory disease (PID), ffbroid, and ovarian cyst with epithelial ovarian cancer. Logistic regression was used to calculate odds ratios (OR) and 95% conffdence intervals (CI) for the associations between individual and composite gynecologic conditions and ovarian cancer.

Results 600 cases and 752 controls enrolled in the African American Cancer Epidemiology Study between 1 December 2010 and 31 December 2015 comprised the study population. After adjusting for potential confounders, a history of endometriosis was associated with ovarian cancer (OR 1.78; 95% CI 1.09–2.90). A non-signiffcant association of similar magnitude was observed with PID (OR 1.33; 95% CI 0.82–2.16), while no association was observed in women with a history of ffbroid or ovarian cyst. A positive trend was observed for an increasing number of reported gynecologic conditions ($p = 0.006$) with consistency across histologic subtypes and among both oral contraceptive users and non-users.

Conclusion A self-reported history of endometriosis among African-American women was associated with increased risk of ovarian cancer. Having multiple benign gynecologic conditions also increased ovarian cancer risk.

Keywords Ovarian cancer · African-American · Endometriosis · Pelvic inffammatory disease (PID) · Ovarian cyst · Uterine ffbroid · African-American Cancer Epidemiology Study (AACES)

Abbreviations

PID	Pelvic inffammatory disease
OC	Oral contraceptive
AACES	African-American Cancer Epidemiology Study
SEER	Surveillance, Epidemiology, and End Results
AJCC	American Joint Committee on Cancer
OR	Odds ratio
CI	Conffdence interval
BMI	Body mass index

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10552-018-1082-4) contains supplementary material, which is available to authorized users.

✉ Hyo K. Park
hpar@med.wayne.edu

Extended author information available on the last page of the article

Introduction

Accumulating epidemiologic evidence suggests that endometriosis is associated with approximately twofold increased risk of developing non-serous epithelial ovarian cancer [1–4]. Studying the pathophysiology and biologic risk factors associated with endometriosis has helped elucidate potential mechanisms of tumorigenesis in non-serous ovarian cancer subtypes distinct from that of serous carcinoma. Chronic inffammation, aberrant immune response, genetic alterations, and hormonal imbalance marked by excess estrogen have been implicated in the multi-step malignant transformation of benign endometriotic cells [5–8]. The epidemiologic linkage between endometriosis and ovarian cancer and the strength of the associations estimated from studies

Published online: 29 September 2018

of predominantly white women remain to be conffrmed in other race and ethnicities.

Other gynecologic conditions, such as pelvic inffammatory disease (PID) [9–11] and ovarian cyst [12], have been associated with increased risk of ovarian cancer in a small number of studies; however, ffndings are conffcting [4, 13–16]. The association between uterine ffbroids, a condition which disproportionately affects African-American women [17, 18], and ovarian cancer is largely unknown. Any potential association observed between ffbroids and ovarian cancer may be modiffed or confounded by increased rates of hysterectomy and procedure-related interruption of tubal patency and ovarian blood supply in women with ffbroids [19–21]. Similarly, oral contraceptive (OC) is frequently prescribed as treatment for benign gynecologic conditions, and OC use could potentially alter the ovarian cancer risk associated with benign gynecologic conditions.

The link between these common benign gynecologic conditions and ovarian cancer remains under-studied in African-Americans. In this study, we explore the relationship between self-reported history of benign gynecologic conditions (endometriosis, PID, uterine ffbroid, and ovarian cyst) and epithelial ovarian cancer in African-American women. While the exact biological etiologies remain to be fully elucidated, these gynecologic pathologies all affect a pro-inffammatory milieu. The association between having multiple gynecologic conditions and ovarian cancer was also examined to assess the potential effect of the increased burden of inffammation-related exposures.

Materials and methods

The data used in these analyses were collected as part of the African-American Cancer Epidemiology Study (AACES), a population-based, case–control study of ovarian cancer in African-American women from 11 geographic regions (Alabama, Georgia, Illinois, Louisiana, Michigan, New Jersey, North Carolina, Ohio, South Carolina, Tennessee, and Texas). Study participants completed informed consent prior to enrollment in the study and institutional review board approval was obtained from all participating institutions. The methods of the study have been previously reported in detail [22], and a brief summary of the study methods follows.

Cases were identified through rapid case ascertainment systems using either state cancer registries, Surveillance, Epidemiology, and End Results (SEER) registries, or individual hospital registries. Inclusion criteria were as follows: self-identiffed African-American/Black race, age 20–79 years at diagnosis, pathology-conffrmed invasive epithelial ovarian cancer diagnosis between 1 December 2010 and 31 December 2015, and ability to complete an interview in English. Controls were identiffed through random digit dialing and frequency matched to cases on 5-year age groups and geographic region. Controls were eligible if they had at least one intact ovary, self-identiffed as African-American/black race, and were 20–79 years at baseline interview. Accrual began in December 2010, and the current analyses include 600 cases and 752 controls enrolled in the study as of December 2017.

Participants were asked to complete a baseline interviewer-administered, computer-assisted telephone survey. Information collected included demographic characteristics; reproductive, gynecologic and medical history; hormone use; family history of cancer; and lifestyle characteristics such as smoking, alcohol consumption, and physical activity. In addition, participants were asked if they had ever been diagnosed with endometriosis, PID, uterine ffbroid or ovarian cyst (yes/no). The interviewer provided a scripted description of the conditions if the participant was not familiar with the medical terminology. If a participant reported a history of these conditions, she was asked to provide the age at ffrst diagnosis. In our analyses, participants who were diagnosed with any gynecologic condition 1 year or less before ovarian cancer diagnosis (cases) or interview date (controls) were coded as not having the condition to reduce surveillance bias. A sensitivity analysis (diagnosis of gynecologic condition 3, 5, or 10 years or less before ovarian cancer diagnosis or baseline interview coded as not having the condition) was performed to evaluate the length of time between diagnosis of gynecologic condition and the referent date (ovarian cancer diagnosis or baseline interview) and its association with ovarian cancer risk.

Overall, 8.7% of cases and 2.5% of controls completed a shorter version of the survey. All variables examined in our analysis were ascertained in both the long and short versions of the survey. Missing data for endometriosis (4 cases), ffbroid (1 cases), PID (5 cases, 2 controls), and ovarian cyst (1 control) were conservatively coded as not having the condition. The distribution of demographic and descriptive characteristics, including frequency of reported gynecologic conditions, between cases and controls was compared using Student's t-test and Chi-square test for continuous and categorical/ordinal variables, respectively. For cases, the mean age at ovarian cancer diagnosis was compared among those with and without a history of each gynecologic condition using Student's *t* test. In addition, the distribution of histologic subtype and American Joint Committee on Cancer (AJCC) stage was summarized by gynecologic condition.

Logistic regression analyses were performed to calculate odds ratios (OR) and 95% conffdence intervals (CI) for the associations between history of endometriosis, PID, uterine ffbroid or ovarian cyst and the risk of ovarian cancer. Known or potential confounders were selected a priori and included in the multivariable model as follows: reference age (age at

diagnosis for cases, age at baseline interview for controls) category (20–29, 30–49, 50–69, 70–79), geographic region (South/mid-Atlantic, South Central, Midwest), marital status (single/never married, married/living with partner, divorced/separated/widowed), education (high school or less, some post-high school training, college or graduate degree), body mass index (BMI in kg/m^2, continuous variable), parity (0, 1, 2, 3 or more), tubal ligation (yes/no), duration of OC use (never, <60 months, ≥ 60 months), ffrst degree family history of breast or ovarian cancer (yes/no), talc use (never use, any genital use, non-genital use only), endometriosis (yes/no), PID (yes/no), ffbroid (yes/no), and ovarian cyst (yes/no). An expanded regression model additionally included hysterectomy status (yes/no) to examine the potential confounding effect of hysterectomy. Hysterectomy status was limited to those performed more than 1 year before the ovarian cancer diagnosis or baseline interview to reduce detection bias.

To explore a potential dose–response relationship, multivariable logistic regression analyses were performed to calculate the association between the total number of benign conditions (0, 1, 2, or more) and risk of ovarian cancer. ORs are reported from categorical models and *p* values for trend are reported from continuous models to test for the linear trend related to an increasing number of benign conditions. The referent group was women with no history endometriosis, PID, ffbroid, or ovarian cyst.

The association between the benign conditions and ovarian cancer risk was further examined in a stratiffed analysis by histologic subtype (serous/non-serous). Non-serous subtypes were further stratiffed into endometrioid, mucinous, clear cell, or other subtype in a supplemental analysis. In addition, the potential modifying effect of OC use on ovarian cancer risk associated with gynecologic conditions was evaluated in a stratiffed analysis by history of OC use (never use/ever use). The interaction between history of OC use and gynecologic conditions was assessed by including a multiplicative term in the models. All statistical analyses were performed using SAS version 9.3 (Cary, North Carolina).

Results

600 cases and 752 controls were included in the analysis. Comparison of demographic and clinical characteristics of cases and controls is presented in Table 1. Cases were older, less likely to be married or living with a partner, and less likely to have post-high school education compared to controls. Cases also were more likely to report having a ffrst degree female relative with breast or ovarian cancer, former smoking, genital talc use, and nulliparity, compared to controls. Cases were less likely to report history of tubal ligation or OC use, but the proportion reporting hysterectomy was similar between the two groups. Cases were more likely to report endometriosis (8.2% vs. 4.4%, $p=0.004$) and PID (7.3% vs. 4.7%, $p=0.037$). There was no difference in the reporting of uterine ffbroid (41.7% vs. 36.6%, $p=0.056$) and ovarian cyst between cases and controls (13.3% vs. 11.2%, $p=0.226$).

The association between benign gynecologic conditions and risk of epithelial ovarian cancer is shown in Table 2. A history of endometriosis was associated with ovarian cancer (OR 1.78; 95% CI 1.09–2.90) after adjusting for age, study site, marital status, education, BMI, parity, tubal ligation, duration of OC use, family history of breast or ovarian cancer, talc use, and history of PID, ffbroid or ovarian cyst. The adjustment variables are all suggested risk factors for ovarian cancer and some are more common in the African American community. For example, talc use is highly prevalent in the African American community and excluding this variable over-estimated the associations in our analysis (data not shown).

An association was observed in women with a history of PID (OR 1.33; 95% CI 0.82–2.16), although the result did not reach statistical signiffcance. While no association was observed in women with a history ffbroid (OR 1.10; 95% CI 0.86–1.40) and ovarian cyst (OR 1.18; 95% CI 0.92–1.52), a positive trend of increasing OR was observed with increasing number of benign gynecologic conditions ($p=0.006$). For women who reported 2 or more gynecological conditions, 31% had PID, 37% had endometriosis, 64% had cysts, and 93% had ffbroids. Direction and magnitude of associations remained essentially unchanged when hysterectomy status was included in the regression model or when the gynecologic diagnosis was censored at 3, 5, and 10 years from the referent date (data not shown).

The relationship between benign gynecologic conditions and epithelial ovarian cancer stratified by serous vs. non-serous histology is shown in Table 3. Endometriosis was associated with a near threefold increase in non-serous ovarian cancer (OR 2.80; 95% CI 1.53–5.10). Odds of serous ovarian cancer was also increased among women with a history of endometriosis, but the association was not significant (OR 1.29; 95% CI 0.71–2.35). Similarly, non-significant associations were observed for PID with both serous (OR 1.65; 95% CI 0.98–2.79) and non-serous (OR 0.90; 95% CI 0.42–1.91) ovarian cancer. No histologic subtype-specific association was observed with history of fibroid, or ovarian cyst. The risk of both serous and non-serous ovarian cancer increased with increasing number of benign gynecologic conditions. A history of 2 or more conditions was associated with a 1.5- to 2-fold increased risk of serous (OR 1.51; 95% CI 1.00–2.29) and non-serous ovarian cancer (OR 2.13; 95% CI 1.32–3.46). Further analysis of non-serous ovarian cancer stratified by histologic subtypes suggested positive associations between endometriosis

Table 1 Demographic and clinical characteristics of ovarian cancer cases and controls in the African American Cancer Epidemiology Study

Characteristics	Total $n = 1,352$ (%)	Cases $n = 600$ (%)	Control $n = 752$ (%)	p value
Age (mean years, range)	56.3 (20–79)	58.1 (20–79)	55.0 (20–79)	< 0.001
BMI (kg/m^2)	32.3 (14.8–78.3)	32.8 (14.8–74.4)	32.0 (15.9–78.3)	0.064
Marital status				0.001
Single, never married	328 (24.3)	144 (24.0)	184 (24.5)	
Married or living with partner	509 (37.6)	197 (32.8)	312 (41.5)	
Divorced/separated or widowed	515 (38.1)	259 (43.2)	256 (34.0)	
Education				0.021
High school or less	550 (40.7)	269 (44.8)	281 (37.4)	
Some post-high school training	358 (26.5)	147 (24.5)	211 (28.1)	
College or graduate degree	444 (32.8)	184 (30.7)	260 (34.6)	
Menstrual status				0.171
Pre/peri-menopause	386 (28.6)	160 (26.7)	226 (30.1)	
Menopause	966 (71.4)	440 (73.3)	526 (69.9)	
Medical history				
Pulmonary disease[a]	220 (16.3)	96 (16.0)	124 (16.5)	0.809
Diabetes	315 (2,336)	137 (22.8)	178 (23.7)	0.718
Cardiac disease[b]	147 (10.9)	64 (10.7)	83 (11.0)	0.828
Hypertension	829 (61.3)	403 (67.2)	426 (56.7)	< 0.001
Anemia	451 (33.3)	236 (39.3)	215 (28.6)	< 0.001
1st degree female relative with breast/ovarian cancer				< 0.001
Yes	292 (21.6)	158 (26.3)	134 (17.8)	
No	1,060 (78.4)	442 (73.7)	618 (82.2)	
Cigarette smoking				< 0.001
Never smoker	769 (56.9)	332 (55.3)	437 (58.1)	
Current smoker	209 (15.5)	61 (10.2)	148 (19.7)	
Former smoker	374 (27.7)	207 (34.5)	167 (22.2)	
Talc use				< 0.001
Never use	578 (42.8)	224 (37.3)	354 (47.1)	
Any genital use	519 (38.4)	264 (44.0)	255 (33.9)	
Non-genital use only	255 (18.9)	112 (18.7)	143 (19.0)	
Parity (# of live births)				0.033
0	207 (15.3)	111 (18.5)	96 (12.8)	
1	251 (18.6)	108 (18.0)	143 (19.0)	
2	345 (25.5)	144 (24.0)	201 (26.7)	
3+	548 (40.6)	236 (39.4)	312 (41.5)	
Tubal ligation				0.060
Yes	513 (37.9)	211 (35.2)	302 (40.2)	
No	839 (62.1)	389 (64.8)	450 (59.8)	
OC use				< 0.001
Never	346 (25.6)	188 (31.3)	158 (21.0)	
< 60 months	574 (42.5)	237 (39.5)	337 (44.8)	
≥ 60 months	432 (32.0)	175 (29.2)	257 (34.2)	
Hysterectomy[c]				0.605
Yes	311 (23.0)	142 (23.7)	169 (22.5)	
No	1,041 (77.0)	458 (76.3)	583 (77.5)	
Benign gynecologic condition[d]				
Endometriosis	82 (6.1)	49 (8.2)	33 (4.4)	0.004
PID	79 (5.8)	44 (7.3)	35 (4.7)	0.037
Fibroid	525 (38.8)	250 (41.7)	275 (36.6)	0.056
Ovarian cyst	164 (12.1)	80 (13.3)	84 (11.2)	0.226

Table 1 (continued)

Characteristics	Total $n = 1{,}352$ (%)	Cases $n = 600$ (%)	Control $n = 752$ (%)	p value
Histology				
High-grade serous		365 (60.8)		
Low-grade serous		17 (2.8)		
Endometrioid		56 (9.3)		
Clear cell		20 (3.3)		
Mucinous		31 (5.2)		
Carcinosarcoma		16 (2.7)		
Other[e]		75 (12.5)		
Missing		20 (3.3)		
Stage				
I/II		188 (31.3)		
III/IV		366 (61.0)		
Unknown		46 (7.7)		

Missing or unknown data: BMI (4 cases, 1 control), parity (1 case)

BMI body mass index, *OC* oral contraceptive, *PID* pelvic inffammatory disease

[a]Include asthma, emphysema, bronchitis

[b]Include angina, congestive heart failure, coronary artery disease

[c]Surgery completed > 1 year before ovarian cancer diagnosis or interview for indications other than ovarian cancer

[d]Diagnosis made > 1 year before ovarian cancer diagnosis or interview

[e]Include mixed, NOS, other invasive epithelial ovarian carcinoma, borderline serous

Table 2 Crude and adjusted odds ratios for the association between epithelial ovarian cancer and benign gynecologic conditions by type and number of condition

Gynecologic conditions	Cases (%)	Control (%)	Crude OR	95% CI	Adjusted OR[a]	95% CI
Type of gynecologic conditions						
Endometriosis						
No	551 (91.8)	719 (95.6)	1.00	Referent	1.00	Referent
Yes	49 (8.2)	33 (4.4)	1.94	1.23–3.05	1.78	1.09–2.90
PID						
No	556 (92.7)	717 (95.4)	1.00	Referent	1.00	Referent
Yes	44 (7.3)	35 (4.7)	1.62	1.03–2.56	1.33	0.82–2.16
Fibroid						
No	350 (58.3)	477 (63.4)	1.00	Referent	1.00	Referent
Yes	250 (41.7)	275 (36.6)	1.24	0.99–1.54	1.10	0.86–1.40
Ovarian cyst						
No	520 (86.7)	668 (88.8)	1.00	Referent	1.00	Referent
Yes	80 (13.3)	84 (11.2)	1.22	0.88–1.70	1.18	0.83–1.69
# of gynecologic conditions						
0	294 (49.0)	420 (55.9)	1.00	Referent	1.00	Referent
1	214 (35.7)	255 (33.9)	1.20	0.95–1.52	1.18	0.92–1.52
2+	92 (15.3)	77 (10.2)	1.71	1.22–2.39	1.66	1.16–2.38
			p trend = 0.002		p trend = 0.006	

Diagnosis made > 1 year before ovarian cancer diagnosis or interview

OR odds ratio, *CI* conffdence interval, *PID* pelvic inffammatory disease, # number

[a]Fully adjusted model—adjusted for age at diagnosis (cases)/interview (control), study site, marital status, education, BMI, parity, tubal ligation, duration of oral contraceptive use, family history of breast or ovarian cancer, talc use, endometriosis, ffbroid, PID, ovarian cyst. OR for # of gynecologic conditions not adjusted for endometriosis, ffbroid, PID, ovarian cyst

Table 3 Crude and adjusted odds ratios for the association between epithelial ovarian cancer and benign gynecologic conditions stratified by histologic subtypes (serous vs. non-serous)

Benign gynecologic condition	Histologic subtype	Cases (%)	Adjusted OR[a]	95% CI
Endometriosis				
No	Serous	362 (94.3)	1.00	Referent
Yes		22 (5.7)	1.29	0.71–2.35
No	Non-serous	169 (86.2)	1.00	Referent
Yes		27 (13.8)	2.80	1.53–5.10
PID				
No	Serous	351 (91.4)	1.00	Referent
Yes		33 (8.6)	1.65	0.98–2.79
No	Non-serous	185 (94.4)	1.00	Referent
Yes		11 (5.6)	0.90	0.42–1.91
Fibroid				
No	Serous	228 (59.4)	1.00	Referent
Yes		156 (40.6)	1.08	0.82–1.43
No	Non-serous	109 (55.6)	1.00	Referent
Yes		87 (44.4)	1.22	0.85–1.75
Ovarian cyst				
No	Serous	335 (87.2)	1.00	Referent
Yes		49 (12.8)	1.16	0.76–1.75
No	Non-serous	167 (85.2)	1.00	Referent
Yes		29 (14.8)	1.13	0.68–1.90
# of gynecologic conditions				
0	Serous	192 (50.0)	1.00	Referent
1		138 (35.9)	1.18	0.89–1.57
2+		54 (14.1)	1.51	1.00-2.29
			p trend = 0.044	
0	Non-serous	91 (46.4)	1.00	Referent
1		67 (34.2)	1.20	0.82–1.75
2+		38 (19.4)	2.13	1.32–3.46
			p trend = 0.004	

Diagnosis made > 1 year before ovarian cancer diagnosis or interview

OR odds ratio, *CI* conffdence interval, *PID* pelvic inffammatory disease

[a]Fully adjusted model—adjusted for age at diagnosis (cases)/interview (control), study site, marital status, education, BMI, parity, tubal ligation, duration of oral contraceptive use, family history of breast or ovarian cancer, talc use, endometriosis, ffbroid, PID, ovarian cyst. OR for # of gynecologic conditions not adjusted for endometriosis, ffbroid, PID, ovarian cyst

and endometrioid (OR 5.17; 95% CI 2.30–11.64) and ovarian cysts with mucinous subtype (OR 3.35; 95% CI 1.33–8.44) (Table S1).

In analyses stratified by history of OC use, there was no consistent pattern or evidence of strong effect modification by OC use on the association between benign gynecologic conditions and ovarian cancer risk (Table 4). The association between endometriosis and ovarian cancer was more pronounced among OC ever- vs. never-users (OR 1.92; 95% CI 1.13–3.24 vs. OR 1.44; 95% CI 0.34–6.31). However, for PID, fibroid, ovarian cyst, and a history of 2 or more benign conditions, the trend was reversed. Test of interaction was not significant for any gynecologic condition.

Discussion

In this analysis of a large, population-based case–control study of African-American women, a history of at least one benign gynecologic condition was reported by approximately half of cases and controls. We observed a consistent association between a history of endometriosis and epithelial ovarian cancer. A consistently positive but non-signiffcant association was observed with PID, while no apparent association was observed with ffbroid or ovarian cyst. Having multiple conditions consistently showed a trend towards increased risk of ovarian cancer across histologic subtypes.

Table 4 Crude and adjusted odds ratios for the association between epithelial ovarian cancer and benign gynecologic conditions stratified by oral contraceptive use

Benign gynecologic condition	Oral contraceptive use	Cases (%)	Control (%)	Adjusted OR[a]	95% CI	$p_{\text{interaction}}$
Endometriosis						0.450
No	OC never use	180 (95.7)	155 (98.1)	1.00	Referent	
Yes		8 (4.3)	3 (1.9)	1.45	0.34–6.31	
No	OC ever use	371 (90.0)	564 (95.0)	1.00	Referent	
Yes		41 (10.0)	30 (5.1)	1.92	1.13–3.24	
PID						0.197
No	OC never use	176 (93.6)	153 (96.8)	1.00	Referent	
Yes		12 (6.4)	5 (3.2)	1.87	0.59–5.95	
No	OC ever use	380 (92.2)	564 (95.0)	1.00	Referent	
Yes		32 (7.8)	30 (5.1)	1.31	0.76–2.26	
Fibroid						0.703
No	OC never use	118 (62.8)	116 (73.4)	1.00	Referent	
Yes		70 (37.2)	42 (26.6)	1.23	0.73–2.06	
No	OC ever use	232 (56.3)	361 (60.8)	1.00	Referent	
Yes		180 (43.7)	233 (39.2)	1.06	0.80–1.40	
Ovarian cyst						0.127
No	OC never use	160 (85.1)	146 (92.4)	1.00	Referent	
Yes		28 (14.9)	12 (7.6)	1.88	0.84–4.20	
No	OC ever use	360 (87.4)	522 (87.9)	1.00	Referent	
Yes		52 (12.6)	72 (12.1)	1.00	0.66–1.51	
# of gynecologic conditions						0.483
0	OC never use	104 (55.3)	108 (68.4)	1.00	Referent	
1		57 (30.3)	39 (24.7)	1.38	0.81–2.33	
2+		27 (14.4)	11 (7.0)	2.36	1.07–5.19	
				p trend = 0.024		
0	OC ever use	190 (46.1)	312 (52.5)	1.00	Referent	
1		157 (38.1)	216 (36.4)	1.12	0.84–1.50	
2+		65 (15.8)	66 (11.1)	1.53	1.01–2.30	
				p trend = 0.055		

Diagnosis made > 1 year before ovarian cancer diagnosis or interview

OR odds ratio, *CI* conffdence interval, *dz.* disease, *PID* pelvic inffammatory disease

[a]Fully adjusted model—adjusted for age at diagnosis (cases)/interview (control), study site, marital status, education, BMI, parity, tubal ligation, family history of breast or ovarian cancer, talc use, endometriosis, ffbroid, PID, ovarian cyst. OR for # of gynecologic conditions not adjusted for endometriosis, ffbroid, PID, ovarian cyst

The most consistent association in our study was observed in women with a history of endometriosis, with increased risk seen across multiple analyses despite the relatively small number of women with the condition. Positive associations between endometriosis and clear cell and endometrioid subtypes conffrm ffndings previously reported in population-based studies of primarily white women [1–4]. The risk of ovarian cancer in women with endometriosis may vary depending on diagnostic criteria used (clinical only vs. surgical-pathological conffrmation), but approximate twofold increased risk observed in our study is consistent with ffndings from the majority of studies examining women with self-reported history of endometriosis (OR 1.3–1.9) [1, 4, 23–26]. Women with a history of endometriosis also had higher odds of being diagnosed with serous ovarian cancer, but the association was not signiffcant. Association between endometriosis and serous ovarian cancer has not been established in existing studies. A recent pooled analysis by Pearce et al. was the ffrst to separately examine the association with high- vs. low-grade serous ovarian cancer and to report a positive association with only low-grade serous subtype [1]. Small sample size in our study precluded further stratiffcation by tumor grade.

Despite the well-established epidemiologic linkage, underlying biological mechanisms driving the association between endometriosis and non-serous ovarian cancer remain to be fully elucidated. Histologically, increased rates of severe atypia with or without complex hyperplasia has

been observed in endometriotic implants adjacent to ovarian carcinoma [2, 6]. This suggests a possible multi-step transformation from benign endometriotic cells to carcinoma aided by the pro-inflammatory microenvironment, altered immune response, and hormonal imbalance. Molecular and genetic studies examining the association between endometriosis and ovarian cancer support the association [7].

We consistently observed an approximate 1.5-fold (up to 1.8-fold among OC never users) increase in ovarian cancer risk among women with a history of PID suggesting a modest association. Observed associations were not consistently significant, but this may be attributed to limitations in sample size and smaller effect size. A small number of case–control and cohort studies have found a 1.5- to twofold increased risk of ovarian cancer in women with a history of PID [9–11], but other studies have reported conflicting results [4, 13, 14]. A recent large pooled analysis of 13 population-based case–control studies found no association between PID and overall ovarian cancer risk, but reported increased risks of low-grade serous and endometrioid subtypes [23]. In our histologic subtype analyses, we observed a positive association with clear cell subtype, but not with endometrioid subtype. Possible linkage with low-grade serous, endometrioid and clear cell subtypes may suggest a shared pro-inflammatory pathway with endometriosis. Supplemental histologic subtype analysis was limited in sample size and exploratory in nature. These results must be interpreted with caution and await further confirmation.

We did not find associations between overall ovarian cancer and a history of fibroid or ovarian cyst, but increasing number of gynecologic conditions was consistently associated with increased risk of ovarian cancer, including both serous and non-serous subtypes. The risk associated with serous ovarian cancer in women with a history of multiple conditions was higher than individual associations observed in any one gynecologic condition. This observation may suggest a possible additive or synergistic effect on tumorigenesis influenced by the pro-inflammatory milieu from an increased burden in the number of benign conditions. Increased risk of serous ovarian cancer in women with other pro-inflammatory risk factors has been reported, most notably in talc users [4, 24].

Direction and magnitude of association and underlying biological mechanism contributing to ovarian cancer tumorigenesis are likely to vary by type of ovarian cyst pathology. Ovarian cyst can represent a wide range of pathologies from functional cysts to benign tumors to endometriomas, which are a type of endometriosis. Existing results vary widely from minimal to no ovarian cancer risk associated with symptomatic functional or stable simple ovarian cyst to twofold or greater increased risk if concomitant infertility or endometrioma is present [15, 16, 25, 26]. An association between ovarian cyst and mucinous ovarian cancer was observed in our histologic subtype analysis. The association between a history of ovarian cyst and mucinous ovarian cancer has not been previously reported, but the linkage is biologically plausible. Positive associations between self-reported history of ovarian cyst and mucinous borderline tumor, believed to be a precursor of invasive mucinous carcinoma, have been reported [12, 16]. More studies are needed to identify the epidemiologic risk factors for mucinous carcinoma, which appear to have molecular and genetic underpinnings distinct from other non-serous subtypes.

Overall, a history of OC use was common among both cases and controls, especially among women with gynecologic conditions. The well-established protective effect of OC has been hypothesized to be mediated by ovulation suppression, reduction in gonadotropins, and increase in apoptosis induced by increased progestin level [27, 28]. In the presence of gynecologic disease, OC may further help modulate ovarian cancer development by preventing hormonal stimulation of endometriotic cells, fibroid, and ovarian cyst and reducing the risk of recurrent PID. We explored the effect of OC use on gynecologic condition-related ovarian cancer risk in a stratified analysis. Overall, OC use did not appear to have a strong or consistent influence on the pattern of associations between benign gynecologic conditions and ovarian cancer beyond the known general protective effect.

This study has limitations that should be considered when interpreting the findings. The prevalence of the gynecologic conditions was based on unverified self-report and subject to misclassification and recall bias. The misclassification may be compounded by the relatively subjective nature of endometriosis or PID diagnosis. Additionally, endometrioma represents a type of ovarian cyst arising from endometriosis and may be reported as a history of ovarian cyst alone. As we do not have information on the type of ovarian cyst in our study, we are not able to estimate the prevalence of this misclassification. To reduce the potential surveillance bias, gynecologic conditions diagnosed within 1 year before ovarian cancer diagnosis or interview date were recoded as not having the condition. We cannot exclude the possibility of bias related to increased intensity and duration of surveillance for more severe disease; however, cases were less likely to have had a health check-up within 2 years and a sensitivity analysis censoring gynecologic diagnosis to 3, 5, or 10 years before ovarian cancer diagnosis demonstrated consistent associations. We also acknowledge that bias due to confounding by treatment of gynecologic conditions other than OC may exist. In our study, hysterectomy was not associated with ovarian cancer, nor did it appear to modify the association between benign gynecologic condition and ovarian cancer. The rate of unilateral oophorectomy among women with ovarian cysts was higher among controls (14 of 84) compared to cases (6 of 85), but small numbers did not allow subgroup analysis.

Our results represent ffndings from the largest case–control study of African-American women with ovarian cancer in the U.S. to date. Moreover, unlike reports from secondary analysis of other studies, AACES was speciffcally designed to investigate risk factors associated with ovarian cancer in African-American women. The large number of participants in our study allowed examination of associations between several common gynecologic conditions and ovarian cancer while adjusting for multiple confounders and known risk factors. In particular, talc powder use is highly prevalent in the African-American community and has been found to be associated with increased risk of ovarian cancer in this and other studies [4, 24, 29]. Indeed, regression models excluding talc use over-estimated the associations in our analyses.

In summary, we report positive associations between a self-reported history of endometriosis, and to a lesser degree PID, with ovarian cancer risk in African-American women similar to existing reports among non-African-American populations. Having more than one benign gynecologic condition also increased ovarian cancer risk.

Acknowledgments We would like to acknowledge the AACES interviewers, Christine Bard, LaTonda Briggs, Whitney Franz (North Carolina) and Robin Gold (Detroit). We also acknowledge the individuals responsible for facilitating case ascertainment across the ten sites including: Jennifer Burczyk-Brown (Alabama); Rana Bayakly, Vicki Bennett and Judy Andrews (Georgia); the Louisiana Tumor Registry; Lisa Paddock and Manisha Narang (New Jersey); Diana Slone, Yingli Wolinsky, Steven Waggoner, Anne Heugel, Nancy Fusco, Kelly Ferguson, Peter Rose, Deb Strater, Taryn Ferber, Donna White, Lynn Borzi, Eric Jenison, Nairmeen Haller, Debbie Thomas, Vivian von Gruenigen, Michele McCarroll, Joyce Neading, John Geisler, Stephanie Smiddy, David Cohn, Michele Vaughan, Luis Vaccarello, Elayna Freese, James Pavelka, Pam Plummer, William Nahhas, Ellen Cato, John Moroney, Mark Wysong, Tonia Combs, Marci Bowling, Brandon Fletcher, (Ohio); Susan Bolick, Donna Acosta, Catherine Flanagan (South Carolina); Martin Whiteside (Tennessee) and Georgina Armstrong and the Texas Registry, Cancer Epidemiology and Surveillance Branch, Department of State Health Services.

Author contributions JS, HP, and MC contributed to the conception and design of the study, analysis and interpretation of data, and drafting of the manuscript. JS, AA, JBS, EF, PT, JJR, and AS contributed to the interpretation of the data and critical revision of the manuscript. All authors reviewed and gave approval of the ffnal version to be published.

Funding This study was supported by the National Cancer Institute (R01CA142081). Additional support was provided by the Metropolitan Detroit Cancer Surveillance System with funding from the National Cancer Institute, National Institute of Health, and the Department of Health and Human Services (Contract HHSN261201000028C), and the Epidemiology Research Core, supported in part by the National Cancer Institute (P30CA22453) to the Karmanos Cancer Institute, Wayne State University School of Medicine. The New Jersey State Cancer Registry, Cancer Epidemiology Services, New Jersey Department of Health, is funded by the Surveillance, Epidemiology and End Results (SEER) Program of the National Cancer Institute under contract HHSN261201300021I, the National Program of Cancer Registries (NPCR), Centers for Disease Control and Prevention under grant 5U58DP003931-02 as well as the State of New Jersey and the Rutgers Cancer Institute of New Jersey.

Compliance with ethical standards

Ethics approval and consent to participate The study protocol and questionnaire were approved by the Institutional Review Boards at Duke University Medical Center, Baylor College of Medicine, Case Western Reserve University School of Medicine, Louisiana State University, Robert Wood Johnson Medical School/Rutgers Cancer Institute, Wayne State University, the University of Alabama-Birmingham, the Medical University of South Carolina, and the University of Tennessee-Knoxville. Additionally, the protocol was approved by central cancer registries in the states of Alabama, Georgia, North Carolina, South Carolina, Tennessee, and Texas, SEER registries in New Jersey, Louisiana, and the Detroit metropolitan area, and 9 individual hospital systems in Ohio. All study participants completed informed consent prior to enrollment.

Availability of data and materials The dataset used and analyzed in this study is available after review from the AACES study investigators and with proper IRB approvals.

Conflict of interest The authors declare that they have no competing interests.

References

1. Pearce CL, Templeman C, Rossing MA, Lee A, Near AM, Webb PM et al (2012) Association between endometriosis and risk of histological subtypes of ovarian cancer: a pooled analysis of case–control studies. Lancet Oncol 13:385–394
2. Somigliana E, Vigano P, Parazzini F, Stoppelli S, Giambattista E, Vercellini P (2006) Association between endometriosis and cancer: a comprehensive review and a critical analysis of clinical and epidemiological evidence. Gynecol Oncol 101:331–341
3. Brinton LA, Sakoda LC, Sherman ME, Frederiksen K, Kjaer SK, Graubard BI et al (2005) Relationship of benign gynecologic diseases to subsequent risk of ovarian and uterine tumors. Cancer Epidemiol Biomark Prev 14:2929–2935
4. Merritt MA, Green AC, Nagle CM, Webb PM (2008) Talcum powder, chronic pelvic inffammation and NSAIDs in relation to risk of epithelial ovarian cancer. Int J Cancer 122:170–176
5. Ness RB (2003) Endometriosis and ovarian cancer: thoughts on shared pathophysiology. Am J Obstet Gynecol 189:280–294
6. Prefumo F, Todeschini F, Fulcheri E, Venturini PL (2002) Epithelial abnormalities in cystic ovarian endometriosis. Gynecol Oncol 84:280–284
7. Lee AW, Templeman C, Stram DA, Beesley J, Tyrer J, Berchuck A et al (2015) Evidence of a genetic link between endometriosis and ovarian cancer. Fertil Steril 105:35–43
8. Wiegand KC, Shah SP, Al-Agha OM, Zhao Y, Tse K, Zeng T et al (2010) ARID1A mutations in endometriosis-associated ovarian carcinomas. N Engl J Med 363:1532–1543
9. Risch HA, Howe GR (1995) Pelvic inffammatory disease and risk of epithelial ovarian cancer. Cancer Epidemiol Biomark Prev 4:447–451
10. Lin HW, Tu YY, Lin SY, Su WJ, Lin WL, Lin WZ et al (2011) Risk of ovarian cancer in women with pelvic inffammatory disease: a population-based study. Lancet Oncol 12:900–904

11. McAlpine JN, Lisonkova S, Joseph KS, McComb PF (2014) Pelvic inflammation and the pathogenesis of ovarian cancer: a cohort study. Int J Gynecol Cancer 24:1406–1413
12. Rossing MA, Cushing-Haugen KL, Wicklung KG, Doherty JA, Weiss NS (2008) Risk of epithelial ovarian cancer in relation to benign ovarian conditions and ovarian surgery. Cancer Causes Control 19:1357–1364
13. Parazzini F, Vecchia CL, Negri E, Moroni S, dal Pino D, Fedele L (1996) Pelvic inflammatory disease and risk of ovarian cancer. Cancer Epidemiol Biomark Prev 5:667–669
14. Rasmussen CB, Faber MT, Jensen A, Hogdall E, Hogdall C, Blaakaer J et al (2013) Pelvic inflammatory disease and risk of invasive ovarian cancer and ovarian borderline tumors. Cancer Causes Control 24:1489–1494
15. Hartge P, Hayes R, Reding D, Sherman ME, Prorok P, Schiffman M et al (2000) Complex ovarian cysts in postmenopausal women are not associated with ovarian cancer risk factors: preliminary data from the prostate, lung, colon, and ovarian cancer screening trial. Am J Obstet Gynecol 183:1232–1237
16. Valentin L, Ameye L, Franchi D, Guerriero S, Jurkovic D, Savelli L et al (2013) Risk of malignancy in unilocular cysts: a study of 1148 adnexal masses classiffed as unilovular cysts at transvaginal ultrasound and review of the literature. Ultrasound Obstet Gynecol 41:80–89
17. Jacoby VL, Fujimoto VY, Giudice LC, Kuppermann M, Washington AE (2010) Racial and ethnic disparities in benign gynecologic conditions and associated surgeries. Am J Obstet Gynecol 202:514–521
18. Wise LA, Palmer JR, Stewart EA, Rosenberg L (2005) Age-speciffc incidence rates for self-reported uterine leiomyomata in the black women's health study. Obstet Gynecol 105:563–568
19. Hankinson SE, Hunter DJ, Colditz GA, Willett WC, Stampfer MJ, Rosner B et al (1993) Tubal ligation, hysterectomy, and risk of ovarian cancer. A prospective study. JAMA 270:2813–2818
20. Rice MS, Hankinson SE, Tworoger SS (2014) Tubal ligation, hysterectomy, unilateral oophorectomy, and risk of ovarian cancer in the Nurses' Health Studies. Fertil Steril 102:192–8
21. Rice MS, Murphy MA, Vitonis AF, Cramer DW, Titus LJ, Tworoger SS et al (2013) Tubal ligation, hysterectomy and epithelial ovarian cancer in the New England Case–control Study. Int J Cancer 133:2415–2421
22. Schildkraut JM, Alberg AJ, Bandera EV, Barnholtz-Sloan J, Bondy M, Cote ML et al (2014) A multi-center population-based case–control study of ovarian cancer in African-American women: the African American Cancer Epidemiology Study (AACES). BMC Cancer 14:688
23. Rasmussen CB, Kjaer SK, Albieri V, Bandera EV, Doherty JA, Hogdall E et al (2017) Pelvic inflammatory disease and the risk of ovarian cancer and borderline ovarian tumors: a pooled analysis of 13 case–control studies. Am J Epidmiol 185:8–20
24. Terry KL, Karageorgi S, Shvetsov YB, Merritt MA, Lurie G, Thompson PJ et al (2013) Genital powder use and risk of ovarian cancer: a pooled analysis of 8,525 cases and 9,859 controls. Cancer Prev Res 6:811–821
25. Ness RB, Cramer DW, Goodman MT, Kjaer SK, Mallin K, Mosgaard BJ et al (2002) Infertility, fertility drugs, and ovarian cancer: a pooled analysis of case–control studies. Am J Epidemiol 155:217–224
26. Lee WL, Chang WH, Wang KC, Guo CY, Chou YJ, Huang N et al (2015) The risk of epithelial ovarian cancer of women with endometriosis may be varied greatly if diagnostic criteria are different. Medicine 94:e1633
27. Schildkraut JM, Calingaert B, Marchbanks PA, Moorman PG, Rodriguez GC (2002) Impact of progestin and estrogen potency in oral contraceptives on ovarian cancer risk. J Natl Cancer Inst 94:32–38
28. Anderson GL, Judd HL, Kaunitz AM, Barad DH, Beresford SA, Pettinger M, Liu J et al (2003) Effects of estrogen plus progestin on gynecologic cancers and associated diagnostic procedures: the Women's Health Initiative randomized trial. JAMA 290:1739–1748
29. Schildkraut JM, Abbot SE, Alberg AJ, Bandera EV, Barnholtz-Sloan JS, Bondy ML et al (2016) Association between body powder use and ovarian cancer: the African American Cancer Epidemiology Study (AACES). Cancer Epidemiol Biomark Prev 25:1411–1417

Affiliations

Hyo K. Park[1] · Joellen M. Schildkraut[2] · Anthony J. Alberg[3] · Elisa V. Bandera[4] · Jill S. Barnholtz-Sloan[5] · Melissa Bondy[6] · Sydnee Crankshaw[7] · Ellen Funkhouser[8] · Patricia G. Moorman[7] · Edward S. Peters[9] · Paul Terry[10] · Frances Wang[7] · Julie J. Ruterbusch[1] · Ann G. Schwartz[1] · Michele L. Cote[1]

Joellen M. Schildkraut
jms2yf@virginia.edu

Anthony J. Alberg
alberg@musc.edu

Elisa V. Bandera
elisa.bandera@rutgers.edu

Jill S. Barnholtz-Sloan
jsb42@case.edu

Melissa Bondy
mbondy@bcm.edu

Sydnee Crankshaw
sydnee.crankshaw@duke.edu

Ellen Funkhouser
efunkhouser@uabmc.edu

Patricia G. Moorman
patricia.moorman@duke.edu

Edward S. Peters
epete1@lsuhsc.edu

Paul Terry
pdterry@ibme.utk.edu

Frances Wang
frances.wang@duke.edu

Julie J. Ruterbusch
ruterbus@karmanos.org

Ann G. Schwartz
schwarta@karmanos.org

Michele L. Cote
cotem@karmanos.org

[1] Department of Oncology and the Karmanos Cancer Institute, Population Studies and Disparities Research Program, Wayne State University School of Medicine, 4100 John R. Mailcode: MM04EP, Detroit, MI 48201, USA

[2] Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA

[3] Hollings Cancer Center and Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA

[4] Department of Population Science, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA

[5] Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH, USA

[6] Cancer Prevention and Population Sciences Program, Baylor College of Medicine, Houston, TX, USA

[7] Department of Community and Family Medicine, Cancer Control and Population Sciences, Duke University School of Medicine, Durham, NC, USA

[8] Division of Preventive Medicine, University of Alabama at Birmingham, Birmingham, AL, USA

[9] Epidemiology Program, Louisiana State University Health Sciences Center School of Public Health, New Orleans, LA, USA

[10] Department of Medicine, University of Tennessee Graduate School of Medicine, Knoxville, TN, USA

Exhibit I

JGIM

REVIEWS

Six Persistent Research Misconceptions

Kenneth J. Rothman, DrPH[1,2]

[1]Research Triangle Institute, Research Triangle Park, NC, USA; [2]Boston University School of Public Health, Boston, MA, USA.

Scientific knowledge changes rapidly, but the concepts and methods of the conduct of research change more slowly. To stimulate discussion of outmoded thinking regarding the conduct of research, I list six misconceptions about research that persist long after their flaws have become apparent. The misconceptions are: 1) There is a hierarchy of study designs; randomized trials provide the greatest validity, followed by cohort studies, with case–control studies being least reliable. 2) An essential element for valid generalization is that the study subjects constitute a representative sample of a target population. 3) If a term that denotes the product of two factors in a regression model is not statistically significant, then there is no biologic interaction between those factors. 4) When categorizing a continuous variable, a reasonable scheme for choosing category cutpoints is to use percentile-defined boundaries, such as quartiles or quintiles of the distribution. 5) One should always report P values or confidence intervals that have been adjusted for multiple comparisons. 6) Significance testing is useful and important for the interpretation of data. These misconceptions have been perpetuated in journals, classrooms and textbooks. They persist because they represent intellectual shortcuts that avoid more thoughtful approaches to research problems. I hope that calling attention to these misconceptions will spark the debates needed to shelve these outmoded ideas for good.

KEY WORDS: study design; data interpretation; epidemiologic methods; representativeness; evaluation of interaction; multiple comparisons; percentile boundaries; statistical significance testing.
J Gen Intern Med 29(7):1060–4
DOI: 10.1007/s11606-013-2755-z
© The Author(s) 2014. This article is published with open access at Springerlink.com

A surprising number of misconceptions persist in the conduct of research involving human subjects. Some persist despite teachings to the contrary, and some because of teachings that should be to the contrary. To spark discussion of these issues, I list here six persistent research misconceptions, and offer a capsule summary of the problems with each of them.

Received November 01, 2013
Revised November 27, 2013
Accepted December 18, 2013
Published online January 23, 2014

Misconception 1. *There is a hierarchy of study designs; randomized trials provide the greatest validity, followed by cohort studies, with case–control studies being least reliable.*

Randomized trials, though often considered the "gold standard" of study types, are not perfect, even in concept. Furthermore, the premise that the comparative validity of study results can be inferred from the type of study is wrong.

Although some believe that evidence from a randomized trial is as compelling as a logical proof, no empirical finding can provide absolute certainty. If randomized trials were perfect, how could they give divergent results? In fact, they are subject to various errors.[1] Obviously there is random error, as one would expect from a study based on random assignment. But there is also systematic error, or bias. For example, randomized trials are usually analyzed using the "intent to treat" principle, which compares the groups that are initially assigned by randomization, regardless of any subsequent non-adherence. Non-adherence results in underestimation of any treatment effect. This bias is usually considered acceptable because it is outweighed by the advantages achieved by random assignment. Underestimation of effects, however, is not acceptable in a safety trial aimed at uncovering adverse effects of the treatment. Another important source of bias in a randomized trial comes from errors in assessing the outcome, such as undercounting of outcome events. Also, even if randomization provides a balance of risk factors between groups at the start of the trial, with extended follow-up, the study groups may become progressively imbalanced through differential attrition or changes in risk factor distributions. With long-term trials, the benefits of random assignment may therefore fade with time.

In short, trials are far from perfect. Furthermore, both cohort and case–control studies will yield valid results when properly designed and carried out. Therefore, mindlessly ascribing greater validity to a study based on a hierarchy of designs[2,3] is fallacious. For example, the relation between cigarette smoking and lung cancer is well established, based on findings from cohort and case–control studies. The connection was never shown clearly in a randomized trial. It is not easy to assign people randomly to smoke or not smoke; however, when smoking cessation was studied as part of a multi-pronged intervention in the randomized Multiple Risk Factor Intervention Trial,[4] those who were

P2.0059.1

urged to cease smoking actually developed more lung cancer than those who did not receive the cessation encouragement. The results of the trial did not overthrow the findings of the many cohort and case–control studies conducted without randomization. Rather, the discrepancy was ascribed to problems with the trial.

In another high-profile example, results from large cohort studies[5,6] indicated that risk of coronary heart disease was reduced among postmenopausal hormone users, but later results from two randomized trials indicated either no association or an increased risk.[7,8] The reaction in the scientific community and the popular press[9] was to discredit the results from the cohort studies, presuming that they had been refuted by the randomized trials. Many continue to believe that interpretation, but in an elegant reanalysis, Hernan et al.[10] showed that the study populations in the cohort studies and the randomized trials were different, and that the effects of postmenopausal hormone use varied greatly according to age and time since menopause. When studies were restricted to new users of hormones, Hernan et al. showed that differences in the distribution of age and time since menopause could explain all of the apparent discrepancies. Although it is common to ascribe such discrepancies to inherent weaknesses of the nonexperimental studies, it is simplistic to assign validity based on a presumed hierarchy of study types.[11]

Similarly, discrepancies between cohort studies and case–control studies should not be explained away superficially by a presumed validity advantage for cohort studies over case–control studies. Properly designed case–control studies will produce the same results as properly designed cohort studies. When conflicts arise, they could stem from problems in either or both types of study. Although case–control studies have long been disparaged as being backwards versions of cohort studies, starting from disease and tracing back to possible causes, epidemiologists today understand case–control studies to be conceptually identical to cohort studies, apart from an efficiency gain that comes from sampling the denominators rather than conducting a complete census. Indeed, the efficiency gain may allow more resources for exposure assessment or case validation in case–control studies, resulting in less bias than in corresponding cohort studies of the same relation.

Those who view case–control studies as backwards versions of cohort studies sometimes make the false analogy that the controls should closely resemble the cases, except that they lack the case-defining disease. In fact, the control group in a case–control study is intended to be a sample of the population denominator that gives rise to the cases, a substitute for the full denominators obtained in a cohort study. Thus, the control group should resemble the entire study population, rather than the cases.[12,13] When properly designed, case–control studies can achieve the same excellent validity as properly designed cohort studies, whereas a poorly designed trial can be unreliable. The type of study should not be taken as a guide to a study's validity.

Misconception 2. *An essential element of making valid generalizations from a study is that the study subjects constitute a representative sample of a target population.*

This misconception is tied to the view that scientific generalization involves the mechanical extrapolation of results from a sample to its source population. But that describes statistical generalization; scientific generalization is different: it is the process of constructing a correct statement about the way nature works.

Scientific generalization is the ultimate goal of scientific inquiry, but a prerequisite is designing a study that has internal validity, which is enhanced by keeping all disturbing variables constant. When have we heard of animal researchers who seek a statistically representative sample of animals? Instead, their operating principle is nearly the opposite of seeking representativeness. Thus, biologists studying mice prefer to study mice that are homogeneous with respect to genes and environment, and that differ only in respect to the experimentally manipulated variable. Unlike the statistical generalization of opinion polls or survey sampling, which merely calls for extrapolation from sample to source population, scientific generalization proceeds by informed guesses, but only from the secure platform of a valid study. Consequently, studies are stronger if they limit variability of confounding factors, as opposed to seeking representativeness. Doll and Hill[14] studied the mortality of male British physicians in relation to their smoking habits. Their findings were considered broadly generalizable despite the fact that their study population was unrepresentative of the general population of tobacco users with regard to sex, race, ethnicity, social class, nationality and many other variables.

When there is a legitimate question about whether an overall association varies by subgroup of some third variable, such as age or ethnic group, it may be necessary to include people drawn from a broad range of values of that third variable, but even then it is counterproductive for the study population to be representative of the source population for that variable. The goal in that case would be to include study subjects distributed evenly across the range, or in a distribution that enhances overall study efficiency. A sample that is representative of the source population will be suboptimal.[15,16]

Misconception 3. *If a term that denotes the product of two factors in a regression model is not statistically significant, then there is no biologic interaction between those factors.*

"Biologic" is meant here broadly, to encompass biochemical, psychological, behavioral and physical interactions. The

problem is that interaction is usually evaluated through regression models, in which the product term addresses statistical interaction rather than biologic interaction.

Biologic interaction refers to two or more causes acting in the same mechanism, with effects that are mutually dependent. It describes a state of nature. If basic effects are measured as changes in disease risk, synergistic (i.e. positive) biologic interaction is present when the joint effect of two causal factors is more than the sum of their effects acting separately.[17] In contrast, statistical interaction does not describe nature; it describes a mathematical model. It is typically assessed with a product term for two variables in a regression model. Its magnitude depends on the choice of measures and scale of measurement. Statistical interaction implies only that the basic functional form of a specific mathematical model is not an apt description of the relation among variables. Two factors that show biologic interaction may or may not exhibit statistical interaction, depending on the model used.

Product terms in regression models have units that can defy interpretation. If one variable is fat consumption, measured in grams per day, and another variable is pack-years of cigarettes smoked, what is the interpretation of a variable that has units of grams/day multiplied by pack-years? The challenge of interpreting such product term coefficients has fostered a focus on the p value accompanying the coefficient, rather than the magnitude of the coefficient itself. Focusing on the pvalue, or on whether the coefficient of a product term is statistically significant, only worsens the problem of mistaking statistical interaction for biologic interaction (see misconception 6). A more meaningful assessment of interaction would be to focus on the proportion of cases of a disease that one could attribute to biologic interaction.[17,18]

Consider a simple example from the TREAT trial (Trial to Reduce Cardiovascular Events with Aranesp Therapy),[19] which evaluated the risk of stroke among 4,038 patients with diabetes mellitus, chronic kidney disease, and anemia randomized to receive darbepoetin alfa or placebo. Among patients without a history of stroke, the risk of stroke during the study period was 2 % among patients receiving placebo and 4 % among patients receiving darbepoeitin alfa. Among patients with a history of stroke, the corresponding risks were 4 % and 12 %. The authors noted that the risk increase was greater for darbepoeitin alfa among those with a history of stroke, but they dismissed this interaction because the product term in a logistic regression model was not statistically significant. The increased risk attributable to darbepoeitin alfa was 2 % in the patients without a history of stroke and 8 % among patients with a history of stroke, indicating strong biologic interaction between darbepoeitin alfa and history of stroke. If the risks were merely additive, the risk would be 6 % among those with both risk factors, instead of the actual 12 %. Thus, half of the risk among those with both risk factors appears attributable to biologic interaction, despite the authors' claim that there was no interaction.

Misconception 4. *When categorizing a continuous variable, a reasonable scheme for choosing category cut-points is to use percentile-defined boundaries, such as quartiles or quintiles of the distribution.*

There are two reasons why using percentiles is a poor method for choosing category boundaries. First, these boundaries may not correspond to the parts of the distribution where biologically meaningful changes occur. Suppose you were conducting a study of vitamin C intake and scurvy risk in the U.S. If you decided to categorize vitamin C intake by quintiles, you would find that the entire relation between vitamin C consumption and scurvy was confined to the lowest quintile, and within that category, to only a small proportion of people who were outliers in their low vitamin C intake. 10 mg/day of vitamin C can prevent scurvy, but those consuming less than that represent a fraction of 1 % of the population in the U. S.[20] Using percentile-based categories would make it impossible to find the effect of inadequate vitamin C intake on scurvy risk, because all intake above 10 mg/d is essentially equivalent. If we routinely use percentile cut-points, we may not know if we are facing the same problem as we would face in the study of vitamin C and scurvy. A more effective alternative would be to begin with many narrow categories, merging neighboring categories until meaningful breaks in risk become evident.

The second problem with percentile-based categories is the difficulty in comparing results across studies, because categories across studies using percentile category boundaries are unlikely to correspond. This problem can be averted by expressing boundary points in terms of the natural units of the variable (such as mg/d for vitamin C intake). It is also useful to report within-category means or medians.

Misconception 5. *One should always report P values or confidence intervals that have been adjusted for multiple comparisons.*

Traditional adjustments for multiple comparisons involve inflating the P value or the width of a confidence interval according to the number of comparisons conducted. If one is analyzing biological data that are replete with actual associations, the premise for traditional adjustments is shaky and the adjustments are difficult to defend. The concern for multiple comparisons stems from fear of finding falsely significant findings (type I errors in the lingo of statistics). In misconception 6, we discuss the problems with using statistical significance testing for data analysis in the first place. But before considering those problems, let us consider the rationale for adjusting reported results for multiple comparisons.

Despite the fact that a single significance test is intended to have a 5 % probability (at the conventionally used level) of being significant when the null hypothesis is true, and

therefore multiple tests when properly carried out should each have this property, there is a concern that when making multiple tests, the probability of a spurious result is increased. Of course, as the number of tests increases, the probability that one or more of them would be falsely positive increases, but that is only because many tests are being conducted. Adjustments for multiple comparisons will reduce these type I errors, but they do so at the expense of increasing type II errors, which are nonsignificant test results in the presence of a real association. When observed associations are all the result of chance, type I errors can occur, but type II errors cannot occur. Conversely, when the observed associations all reflect actual relationships, type II errors can occur, but type I errors cannot. Thus, the context of any analysis has fundamental implications regarding the interpretation of the data. In particular, it is absurd to make adjustments that reduce type I errors at the expense of increasing type II errors without some evaluation of the estimated relative cost and frequency of each type of error.

If scientists were put to work studying random numbers instead of biologic data, all the significant results they reported would represent type I errors, and adjustments for multiple comparisons would make sense; some skeptics believe that studies of genome-wide association scans may approximate this situation.[21] But when scientists are studying biological relations rather than random numbers, the premise that type I errors are the major concern may be wrong.[22] A more rigorous evaluation of the need for multiplicity adjustments would begin with an assessment of the tenability of the thesis that the data are essentially random numbers. If one is studying experiments on psychic phenomena, skepticism about the results might lend support to multiplicity adjustments. If one is studying physiologic effects of pharmaceutical agents, real associations are to be expected and the adjustments are more difficult to defend. Studying single nucleotide polymorphisms in relation to a given disease might be a middle ground. One approach to this issue that is theoretically more defensible is a Bayesian approach, which assigns prior credibility to various levels of association and adjusts by using Bayes' theorem to calculate posterior credibility.[23,24]

Misconception 6. *Significance testing is useful and important for the interpretation of data.*

Significance testing has led to far more misunderstanding and misinterpretation than clarity in interpreting study results.[25–28] A significance test is a degraded version of the P value, a statistic that blends precision with effect size, thus confusing two essential aspects of data interpretation. Measuring effect size and its precision as separate tasks is a more direct and clearer approach to data interpretation.

For research studies that aim to measure associations, and infer whether they reflect causal connections, focusing on the magnitude of these associations ought to be the primary goal: estimation of effects is decidedly preferable to statistical testing. Ideally, a study estimates the magnitude of the effect size, and analyzes the possible errors that might have distorted it. Systematic errors such as confounding from measured factors can be dealt with through analytic methods; other systematic errors, such as the effects of measurement error or selection bias, can be addressed through sensitivity analyses (also known as bias analysis). Random error is typically expressed through confidence intervals, giving a range of parameter values that are consistent with the data to a specified level.

It is unfortunate that a confidence interval, from which both an estimate of effect size and its measurement precision can be drawn, is typically used merely to judge whether it contains the null value or not, thus converting it to a significance test. Significance tests are a poor classification scheme for study results; strong effects may be incorrectly interpreted as null findings because authors fallaciously interpret lack of statistical significance to imply lack of effect, or weak effects may be incorrectly interpreted as important because they are statistically significant. Rather than be used as surrogate significance tests, confidence intervals ought to be interpreted as quantitative measures indicating magnitude of effect size and degree of precision, with little attention paid to the precise location of the boundaries of the confidence interval. This advice is backed by the Uniform Requirements for Manuscripts Submitted to Biomedical Journals, but nevertheless often overlooked even by reviewers and editors whose journals support the requirements.[29]

Many misconceptions derive from reliance on statistical significance testing. The focus on the statistical significance of interaction terms instead of measuring interaction, as discussed above, is one example. The evaluation of dose–response trends simply by declaring that there is or is not a significant trend, rather than expressing the magnitude and ideally the shape of that trend, is another. Yet another is the advice sometimes offered to calculate the power of a study when reporting results, especially if those results are not statistically significant. Reporting the power of a study as part of the results is called "post-hoc" power calculation.[30] Power calculations are based on a hypothesis about the level of association that is to be distinguished from a null association, but when the study results are on hand, there is no longer any need to hypothesize about the magnitude of the association, because you now have an estimate of it. A confidence interval for the estimated association conveys all the relevant information; nothing further is to be gained from a power calculation.

The unfortunate consequence of the focus on statistical significance testing has been to foster a dichotomous view of relationships that are better assessed in quantitative terms. This distinction is more than a nicety. Every day there are important, regrettable and avoidable misinterpretations of data that results from the confusing fog of

statistical significance testing. Most of these errors could be avoided if the focus were shifted from statistical testing to estimation.

CONCLUSION

Why do such important misconceptions about research persist? To a large extent these misconceptions represent substitutes for more thoughtful and difficult tasks. It is simpler to resolve a discrepancy between a trial and a nonexperimental study in favor of the trial, without undertaking the laborious analysis that Hernan et al. did.[10] It is easy to declare that a result is not statistically significant, falsely implying that there is no indication of an association, rather than to consider quantitatively the range of associations that the data actually support. These misconceptions involve taking the low road, but when that road is crowded with others taking the same path, there may be little reason to question the route. Indeed, these misconceptions are often perpetuated in journals, classrooms and textbooks. I believe that the best prospect for improvement is to raise consciousness about the issues, with reasoned debate. Max Planck once said, “A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it.”[31] To the extent that this cynical view is correct, we can expect to see outmoded concepts fade away slowly at best. I hope that calling attention to these misconceptions will spark the needed debates and be a catalyst for change.

Acknowledgements: *I received helpful criticism from Susana Perez, Andrea Margulis, Manel Pladevall, and Jordi Castellsague.*

Conflict of Interest: *The author declares no conflict of interest.*

Corresponding Author: *Kenneth J. Rothman, DrPH; Research Triangle Institute, Research Triangle Park, NC, USA (e-mail: KRothman@rti.org).*

Open Access *This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.*

REFERENCES

1. **Hernán MA, Hernández-Díaz S, Robins JM.** Randomized trials analyzed as observational studies. Ann Intern Med. 2013;159:560–2. doi:10.7326/0003-4819-159-8-201310150-00709
2. **Ioannidis JPA.** Why most published research findings are false. PLoS Med. 2005;2(8):e124.
3. **Hiatt WR.** Observational studies of drug safety–aprotinin and the absence of transparency. N Engl J Med. 2006;355:2171–3.
4. **Shaten BJ, Kuller LH, Kjelsberg MO, Stamler J, Ockene JK, Cutler JA, Cohen JD.** Lung cancer mortality after 16 years in MRFIT participants in intervention and usual-care groups. Multiple Risk Factor Intervention Trial. Ann Epidemiol 1997;7:125–36.
5. **Grodstein F, Manson JE, Colditz GA, et al.** A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease. Ann Intern Med. 2000;133:933–41.
6. **Varas-Lorenzo C, García-Rodríguez LA, Pérez-Gutthann S, et al.** Hormone replacement therapy and incidence of acute myocardial infarction. Circulation. 2000;101:2572–8.
7. **Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B, Vittinghoff E.** Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. JAMA. 1998;280:605–13. doi:10.1001/jama.280.7.605.
8. **Manson JE, Hsia J, Johnson KC, et al.** Estrogen plus progestin and the risk of coronary heart disease. N Engl J Med. 2003;349:523–34.
9. **Taubes G.** Do we really know what makes us healthy? New York Times, September 16, 2007.
10. **Hernán MA, Alonso A, Logan R, Grodstein F, Michels K, Willett WC, Manson JE, Robins JM.** Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology. 2008;19:766–79. doi:10.1097/EDE.0b013e3181875e61
11. **Concato J.** Observational versus experimental studies: what’s the evidence for a hierarchy? NeuroRx. 2004;1:341–7.
12. **Vandenbroucke JP, Pearce N.** Case–control studies: basic concepts. Int J Epidemiol. 2012;41:1480–9. doi:10.1093/ije/dys147.
13. **Rothman KJ.** Chapter 5, Types of Epidemiologic Studies, in *Epidemiology, An Introduction*, 2nd Edition. Oxford University Press, New York, 2012.
14. **Doll R, Hill AB.** The mortality of doctors in relation to their smoking habits: a preliminary report. Br Med J. 1954;ii:1451–5.
15. **Rothman KJ, Gallacher J, Hatch EE.** Why representativeness should be avoided. Int J Epidemiol. 2013;42:1012–4. doi:10.1093/ije/dys223
16. **Rothman KJ, Gallacher J, Hatch EE.** When it comes to scientific inference, sometimes a cigar is just a cigar. Int J Epidemiol. 2013;42:1026–8. doi:10.1093/ije/dyt124
17. **Rothman KJ.** Chapter 11, Measuring Interaction, in *Epidemiology, An Introduction*, 2nd Edition. Oxford University Press, New York, 2012.
18. **Knol MJ, van der Tweel I, Grobbee DE, Numans ME, Geerlings MI.** Estimating interaction on an additive scale between continuous determinants in a logistic regression model. Int J Epidemiol. 2007;36:1111–8.
19. **Skali H, Parving HH, Parfrey PS, Burdmann EA, Lewis EF, Ivanovich P, Keithi-Reddy SR, McGill JB, McMurray JJ, Singh AK, Solomon SD, Uno H, Pfeffer MA.** TREAT Investigators: Stroke in patients with type 2 diabetes mellitus, chronic kidney disease, and anemia treated with Darbepoetin Alfa: the trial to reduce cardiovascular events with Aranesp therapy (TREAT) experience. Circulation. 2011;124:2903–8.
20. Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids. Institute of Medicine, The National Academies Press, Washington, D. C., 2000.
21. **Dudbridge F, Gusnanto A.** Estimation of significance thresholds for genomewide association scans. Genet Epidemiol. 2008;32:227–34.
22. **Rothman KJ.** No adjustments are needed for multiple comparisons. Epidemiology. 1990;1:43–6.
23. **Greenland S, Robins J.** Empirical-Bayes adjustments for multiple comparisons are sometimes useful. Epidemiology. 1991;2:244–51.
24. **Greenland S, Poole C.** Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. Arch Environ Health. 1994;48:9–16.
25. **Rothman KJ.** A show of confidence (editorial). N Engl J Med. 1978;299:1362–3.
26. **Poole C.** Beyond the confidence interval. Am J Public Hlth. 1987;77:195–9.
27. **Rothman KJ.** Significance questing (Editorial). Ann Int Med. 1986;105:445–7.
28. **Gelman A, Stern H.** The difference between “Significant” and “Not Significant” is not itself statistically significant. Amer Statistician. 2006;60:328–31.
29. Uniform Requirements for Manuscripts Submitted to Biomedical Journals, http://www.icmje.org/manuscript_1prepare.html (accessed May 2, 2013)
30. **Smith AH, Bates MN.** Confidence limit analyses should replace power calculations in the interpretation of epidemiologic studies. Epidemiology. 1992;3:449–52.
31. **Planck M.** *Scientific Autobiography and Other Papers*, Philosophical Library, New York, 1968, trans. F. Gaynor (New York, 1949), pp. 33–34

Copyright of JGIM: Journal of General Internal Medicine is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Exhibit J

COMMENT

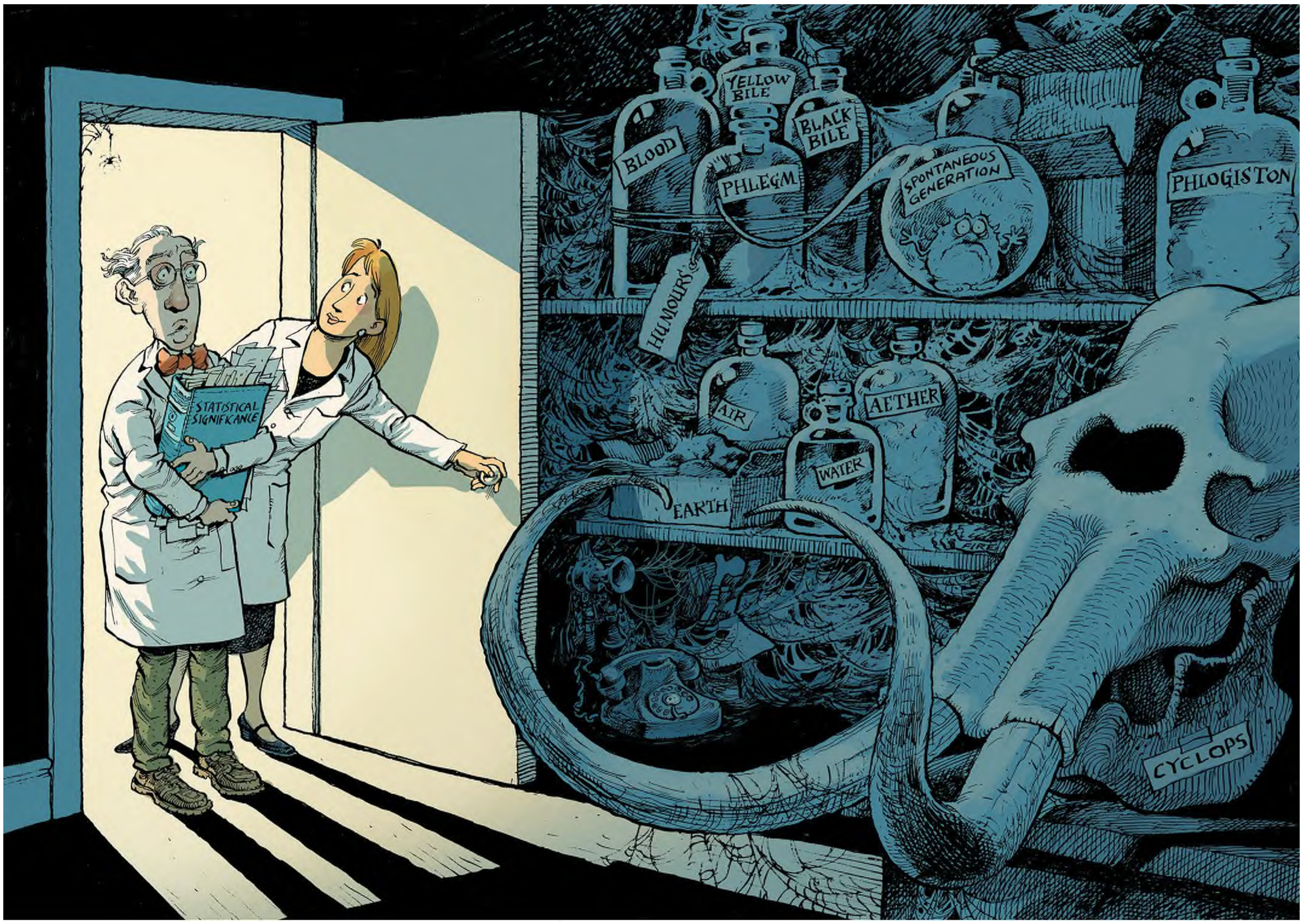
EVOLUTION Cooperation and conflict from ants and chimps to us **p.308**

HISTORY To fight denial, study Galileo and Arendt **p.309**

CHEMISTRY Three more unsung women — of astatine discovery **p.311**

PUBLISHING As well as ORCID ID and English, list authors in their own script **p.311**

ILLUSTRATION BY DAVID PARKINS



Retire statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

When was the last time you heard a seminar speaker claim there was 'no difference' between two groups because the difference was 'statistically non-significant'?

If your experience matches ours, there's a good chance that this happened at the last talk you attended. We hope that at least someone in the audience was perplexed if, as frequently happens, a plot or table showed that there actually was a difference.

How do statistics so often lead scientists to deny differences that those not educated in statistics can plainly see? For several generations, researchers have been warned that a statistically non-significant result does not 'prove' the null hypothesis (the hypothesis that there is no difference between groups or no effect of a treatment on some measured outcome)[1]. Nor do statistically significant results 'prove' some other hypothesis. Such misconceptions have famously warped the literature with overstated claims and, less famously, led to claims of conflicts between studies where none exists.

We have some proposals to keep scientists from falling prey to these misconceptions.

PERVASIVE PROBLEM

Let's be clear about what must stop: we should never conclude there is 'no difference' or 'no association' just because a *P* value is larger than a threshold such as 0.05 ▸

© 2019 Springer Nature Limited. All rights reserved.

▶ or, equivalently, because a confidence interval includes zero. Neither should we conclude that two studies conflict because one had a statistically significant result and the other did not. These errors waste research efforts and misinform policy decisions.

For example, consider a series of analyses of unintended effects of anti-inflammatory drugs[2]. Because their results were statistically non-significant, one set of researchers concluded that exposure to the drugs was "not associated" with new-onset atrial fibrillation (the most common disturbance to heart rhythm) and that the results stood in contrast to those from an earlier study with a statistically significant outcome.

Now, let's look at the actual data. The researchers describing their statistically non-significant results found a risk ratio of 1.2 (that is, a 20% greater risk in exposed patients relative to unexposed ones). They also found a 95% confidence interval that spanned everything from a trifling risk decrease of 3% to a considerable risk increase of 48% ($P = 0.091$; our calculation). The researchers from the earlier, statistically significant, study found the exact same risk ratio of 1.2. That study was simply more precise, with an interval spanning from 9% to 33% greater risk ($P = 0.0003$; our calculation).

It is ludicrous to conclude that the statistically non-significant results showed "no association", when the interval estimate included serious risk increases; it is equally absurd to claim these results were in contrast with the earlier results showing an identical observed effect. Yet these common practices show how reliance on thresholds of statistical significance can mislead us (see 'Beware false conclusions').

These and similar errors are widespread. Surveys of hundreds of articles have found that statistically non-significant results are interpreted as indicating 'no difference' or 'no effect' in around half (see 'Wrong interpretations' and Supplementary Information).

In 2016, the American Statistical Association released a statement in *The American Statistician* warning against the misuse of statistical significance and *P* values. The issue also included many commentaries on the subject. This month, a special issue in the same journal attempts to push these reforms further. It presents more than 40 papers on 'Statistical inference in the 21st century: a world beyond $P < 0.05$'. The editors introduce the collection with the caution "don't say 'statistically significant'"[3]. Another article[4] with dozens of signatories also calls on authors and journal editors to disavow those terms.

We agree, and call for the entire concept of statistical significance to be abandoned.

"Eradicating categorization will help to halt overconfident claims, unwarranted declarations of 'no difference' and absurd statements about 'replication failure'."

We are far from alone. When we invited others to read a draft of this comment and sign their names if they concurred with our message, 250 did so within the first 24 hours. A week later, we had more than 800 signatories — all checked for an academic affiliation or other indication of present or past work in a field that depends on statistical modelling (see the list and final count of signatories in the Supplementary Information). These include statisticians, clinical and medical researchers, biologists and psychologists from more than 50 countries and across all continents except Antarctica. One advocate called it a "surgical strike against thoughtless testing of statistical significance" and "an opportunity to register your voice in favour of better scientific practices".

We are not calling for a ban on *P* values. Nor are we saying they cannot be used as a decision criterion in certain specialized applications (such as determining whether a manufacturing process meets some quality-control standard). And we are also not advocating for an anything-goes situation, in which weak evidence suddenly becomes credible. Rather, and in line with many others over the decades, we are calling for a stop to the use of *P* values in the conventional, dichotomous way — to decide whether a result refutes or supports a scientific hypothesis[5].

QUIT CATEGORIZING

The trouble is human and cognitive more than it is statistical: bucketing results into 'statistically significant' and 'statistically non-significant' makes people think that the items assigned in that way are categorically different[6–8]. The same problems are likely to arise under any proposed statistical alternative that involves dichotomization, whether frequentist, Bayesian or otherwise.

Unfortunately, the false belief that crossing the threshold of statistical significance is enough to show that a result is 'real' has led scientists and journal editors to privilege such results, thereby distorting the literature. Statistically significant estimates are biased upwards in magnitude and potentially to a large degree, whereas statistically non-significant estimates are biased downwards in magnitude. Consequently, any discussion that focuses on estimates chosen for their significance will be biased. On top of this, the rigid focus on statistical significance encourages researchers to choose data and methods that yield statistical significance for some desired (or simply publishable) result, or that yield statistical non-significance for an undesired result, such as potential side effects of drugs — thereby invalidating conclusions.

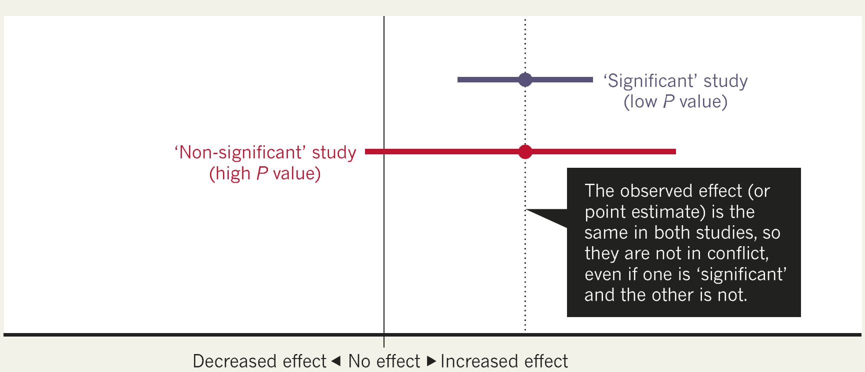
The pre-registration of studies and a commitment to publish all results of all analyses can do much to mitigate these issues. However, even results from pre-registered studies can be biased by decisions invariably left open in the analysis plan[9]. This occurs even with the best of intentions.

Again, we are not advocating a ban on *P* values, confidence intervals or other statistical measures — only that we should not treat them categorically. This includes dichotomization as statistically significant or not, as well as categorization based on other statistical measures such as Bayes factors.

One reason to avoid such 'dichotomania' is that all statistics, including *P* values and confidence intervals, naturally vary from study to study, and often do so to a surprising degree. In fact, random variation alone can easily lead to large disparities in *P* values, far beyond falling just to either side of the 0.05 threshold. For example, even if researchers could conduct two perfect replication studies of some genuine effect, each with 80% power (chance) of achieving $P < 0.05$, it would not be very surprising for one to obtain $P < 0.01$ and the other $P > 0.30$.

BEWARE FALSE CONCLUSIONS

Studies currently dubbed 'statistically significant' and 'statistically non-significant' need not be contradictory, and such designations might cause genuine effects to be dismissed.



SOURCE: V. AMRHEIN *ET AL.*

© 2019 Springer Nature Limited. All rights reserved.

Whether a P value is small or large, caution is warranted.

We must learn to embrace uncertainty. One practical way to do so is to rename confidence intervals as 'compatibility intervals' and interpret them in a way that avoids overconfidence. Specifically, we recommend that authors describe the practical implications of all values inside the interval, especially the observed effect (or point estimate) and the limits. In doing so, they should remember that all the values between the interval's limits are reasonably compatible with the data, given the statistical assumptions used to compute the interval[7,10]. Therefore, singling out one particular value (such as the null value) in the interval as 'shown' makes no sense.

We're frankly sick of seeing such nonsensical 'proofs of the null' and claims of non-association in presentations, research articles, reviews and instructional materials. An interval that contains the null value will often also contain non-null values of high practical importance. That said, if you deem all of the values inside the interval to be practically unimportant, you might then be able to say something like 'our results are most compatible with no important effect'.

When talking about compatibility intervals, bear in mind four things. First, just because the interval gives the values most compatible with the data, given the assumptions, it doesn't mean values outside it are incompatible; they are just less compatible. In fact, values just outside the interval do not differ substantively from those just inside the interval. It is thus wrong to claim that an interval shows all possible values.

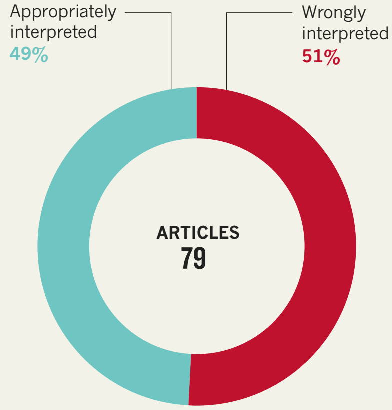
Second, not all values inside are equally compatible with the data, given the assumptions. The point estimate is the most compatible, and values near it are more compatible than those near the limits. This is why we urge authors to discuss the point estimate, even when they have a large P value or a wide interval, as well as discussing the limits of that interval. For example, the authors above could have written: 'Like a previous study, our results suggest a 20% increase in risk of new-onset atrial fibrillation in patients given the anti-inflammatory drugs. Nonetheless, a risk difference ranging from a 3% decrease, a small negative association, to a 48% increase, a substantial positive association, is also reasonably compatible with our data, given our assumptions.' Interpreting the point estimate, while acknowledging its uncertainty, will keep you from making false declarations of 'no difference', and from making overconfident claims.

Third, like the 0.05 threshold from which it came, the default 95% used to compute intervals is itself an arbitrary convention. It is based on the false idea that there is a 95% chance that the computed interval itself contains the true value, coupled with the vague

SOURCE: V. AMRHEIN *ET AL.*

WRONG INTERPRETATIONS

An analysis of 791 articles across 5 journals* found that around half mistakenly assume non-signi cance means no e ect.



*Data taken from: P. Schatz *et al. Arch. Clin. Neuropsychol.* **20**, 1053–1059 (2005); F. Fidler *et al. Conserv. Biol.* **20**, 1539–1544 (2006); R. Hoekstra *et al. Psychon. Bull. Rev.* **13**, 1033–1037 (2006); F. Bernardi *et al. Eur. Sociol. Rev.* **33**, 1–15 (2017).

feeling that this is a basis for a confident decision. A different level can be justified, depending on the application. And, as in the anti-inflammatory-drugs example, interval estimates can perpetuate the problems of statistical significance when the dichotomization they impose is treated as a scientific standard.

Last, and most important of all, be humble: compatibility assessments hinge on the correctness of the statistical assumptions used to compute the interval. In practice, these assumptions are at best subject to considerable uncertainty[7,8,10]. Make these assumptions as clear as possible and test the ones you can, for example by plotting your data and by fitting alternative models, and then reporting all results.

Whatever the statistics show, it is fine to suggest reasons for your results, but discuss a range of potential explanations, not just favoured ones. Inferences should be scientific, and that goes far beyond the merely statistical. Factors such as background evidence, study design, data quality and understanding of underlying mechanisms are often more important than statistical measures such as P values or intervals.

The objection we hear most against retiring statistical significance is that it is needed to make yes-or-no decisions. But for the choices often required in regulatory, policy and business environments, decisions based on the costs, benefits and likelihoods of all potential consequences always beat those made based solely on statistical significance. Moreover, for decisions about whether to pursue a research idea further, there is no simple connection between a P value and the probable results of subsequent studies.

What will retiring statistical significance look like? We hope that methods sections and data tabulation will be more detailed and nuanced. Authors will emphasize their estimates and the uncertainty in them — for example, by explicitly discussing the lower and upper limits of their intervals. They will not rely on significance tests. When P values are reported, they will be given with sensible precision (for example, $P = 0.021$ or $P = 0.13$) — without adornments such as stars or letters to denote statistical significance and not as binary inequalities ($P < 0.05$ or $P > 0.05$). Decisions to interpret or to publish results will not be based on statistical thresholds. People will spend less time with statistical software, and more time thinking.

Our call to retire statistical significance and to use confidence intervals as compatibility intervals is not a panacea. Although it will eliminate many bad practices, it could well introduce new ones. Thus, monitoring the literature for statistical abuses should be an ongoing priority for the scientific community. But eradicating categorization will help to halt overconfident claims, unwarranted declarations of 'no difference' and absurd statements about 'replication failure' when the results from the original and replication studies are highly compatible. The misuse of statistical significance has done much harm to the scientific community and those who rely on scientific advice. P values, intervals and other statistical measures all have their place, but it's time for statistical significance to go. ■

Valentin Amrhein *is a professor of zoology at the University of Basel, Switzerland.* **Sander Greenland** *is a professor of epidemiology and statistics at the University of California, Los Angeles.* **Blake McShane** *is a statistical methodologist and professor of marketing at Northwestern University in Evanston, Illinois. For a full list of co-signatories, see Supplementary Information.*
e-mail: v.amrhein@unibas.ch

1. Fisher, R. A. *Nature* **136**, 474 (1935).
2. Schmidt, M. & Rothman, K. J. *Int. J. Cardiol.* **177**, 1089–1090 (2014).
3. Wasserstein, R. L., Schirm, A. & Lazar, N. A. *Am. Stat.* https://doi.org/10.1080/00031305.2019.1583913 (2019).
4. Hurlbert, S. H., Levine, R. A. & Utts, J. *Am. Stat.* https://doi.org/10.1080/00031305.2018.1543616 (2019).
5. Lehmann, E. L. *Testing Statistical Hypotheses* 2nd edn 70–71 (Springer, 1986).
6. Gigerenzer, G. *Adv. Meth. Pract. Psychol. Sci.* **1**, 198–218 (2018).
7. Greenland, S. *Am. J. Epidemiol.* **186**, 639–645 (2017).
8. McShane, B. B., Gal, D., Gelman, A., Robert, C. & Tackett, J. L. *Am. Stat.* https://doi.org/10.1080/00031305.2018.1527253 (2019).
9. Gelman, A. & Loken, E. *Am. Sci.* **102**, 460–465 (2014).
10. Amrhein, V., Trafimow, D. & Greenland, S. *Am. Stat.* https://doi.org/10.1080/00031305.2018.1543137 (2019).

Supplementary information accompanies this article; see go.nature.com/2tc5nkm

Exhibit K

Fordham Law Review

Volume 52 | Issue 5 Article 2

1984

Epidemiologic Proof in Toxic Tort Litigation

Bert Black

David E. Lilienfeld

Recommended Citation

Bert Black and David E. Lilienfeld, *Epidemiologic Proof in Toxic Tort Litigation*, 52 Fordham L. Rev. 732 (1984).
Available at: hTp://ir.lawnet.fordham.edu/Th/vol52/iss5/2

This Article is brought to you for free and open access by FLASH: The Fordham Law Archive of Scholarship and History. It has been accepted for inclusion in Fordham Law Review by an authorized editor of FLASH: The Fordham Law Archive of Scholarship and History. For more information, please contact tmelnick@law.fordham.edu.

EPIDEMIOLOGIC PROOF IN TOXIC TORT LITIGATION

*BERT BLACK**

and

*DAVID E. LILIENFELD***

TABLE OF CONTENTS

---

* Associate, Venable, Baetjer and Howard, Baltimore, Maryland; B.S. 1974, University of Maryland; M.S. 1975, Georgia Institute of Technology; J.D. 1982, Yale Law School.

** Resident Physician, Minnesota Department of Health, Division of Disease Prevention and Control, Minneapolis, Minnesota; B.A. 1978, Johns Hopkins University; M.S. Eng'g 1980, Johns Hopkins University; M.D. 1984, University of Maryland; Certified in epidemiology by American College of Epidemiology.

The authors wish to express their thanks to Cathleen Lippert, for typing the original manuscript, to John S. Nixdorff, librarian at Venable, Baetjer and Howard, for tracking down countless references, and to Julie R. Silverman for helping with editing and proofreading. The authors also wish to thank Dr. Abraham M. Lilienfeld, Robert G. Smith, Ben Finkelstein, Brigid E. Kenney, Joan E. Hartman and Judy F. Black for their many helpful suggestions.

INTRODUCTION

TOXIC tort[1] litigation has emerged as a major social and legal concern,[2] a development that has engendered numerous proposals for legal reform. Many of these reforms would require institutional

---

1. This Article loosely defines toxic tort cases as those in which the plaintiff seeks compensation for harm allegedly caused by exposure to a substance that increases the risk of contracting a serious disease, but does not cause an immediately apparent response. These cases generally involve a period of latency or incubation prior to the onset of the disease. In most cases the increased risk of the disease does not diminish or dissipate, even with the cessation of exposure. The Article discusses exposure to radiation as well as to chemicals, and considers some cases involving drugs because many of the causation issues are similar to those in environmental or occupational cases. It also considers birth defect cases. The vast majority of toxic tort cases, however, are related to cancer and the issue of carcinogenesis, and thus, parts of this Article focus only on cancer and its causes.

2. One commentator notes that "[e]ven without a crystal ball, it is easy to see a wave of cancer litigation on the horizon." Shelton, *Defending Cancer Litigation: The Causation Defense*, For The Defense, January 1982, at 8, 14. Another cites asbestos litigation as indicative of the trend, and points out that "there are more than 15,000 asbestos related cases now pending, and additional cases are being filed at the rate of over 400 each month; it has been estimated that over 30,000 additional suits will be filed in the next 25 years." Olick, *Chapter 11—A Dubious Solution To Massive Toxic Tort Liability*, 18 Forum 361, 361 (1983). Also part of the trend are claims brought by people alleging harm from exposure to dioxin. *See* Long & Hanson, *Dioxin Issue Focuses on Three Major Controversies in U.S.*, Chem. & Eng'g News, June 6, 1983, at 23, 24. One accident involving dioxin at a West Virginia chemical plant has resulted in claims totaling 700 million dollars. Webber, *Dioxin Liability is Huge Problem for Companies, Courts*, Chem. & Eng'g News, June 6,

innovations, such as administrative funds from which claimants could obtain compensation with relatively little evidence of causation.[3] Most, however, would also allow recovery under existing tort theories.[4] Thus, questions about the application of common law principles in evaluating evidence of causation in toxic tort cases remain open.[5]

1983, at 57, 59. For other examples, see Note, *Establishing Causation in Chemical Exposure Cases: The Precursor Symptoms Theory*, 35 Rutgers L. Rev. 163, 164 n.2 (1982) [hereinafter cited as *Precursor Symptoms*].

3. *See, e.g.*, Ginsberg & Weiss, *Common Law Liability for Toxic Torts: A Phantom Remedy*, 9 Hofstra L. Rev. 859, 928-40 (1981); Milhollin, *Long-Term Liability for Environmental Harm*, 41 U. Pitt. L. Rev. 1, 16-25 (1979); Trauberman, *Statutory Reform of "Toxic Torts": Relieving Legal, Scientific and Economic Burdens on the Chemical Victim*, 7 Harv. Envtl. L. Rev. 177, 237, 243 (1983).

Perhaps the best known proposals for changes in the law are those made by the Superfund Section 301(e) Study Group (Study Group), which was appointed pursuant to Section 301(e) of the Comprehensive Environmental Response, Compensation, and Liability Act of 1980, Pub. L. No. 96-510, 94 Stat. 2767 (1980) (codified at 42 U.S.C. § 9601 et seq. (Supp. V 1981)). The Study Group submitted a report to Congress in September 1982 that recommended the creation of rebuttable presumptions of causation to facilitate access to an administrative victim compensation fund. "Superfund Section 301(e) Study Group," 97th Cong., 2d Sess., Injuries and Damages From Hazardous Wastes—Analysis and Improvement of Legal Remedies—Report to Congress in Compliance with Section 301(e) of the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (P.L. 96-510) 213-25 (Comm. Print 1982) [hereinafter cited as 301(e) Study].

4. The only proposal of which the authors are aware that would eliminate tort law in the area of toxic exposure litigation is that of Ginsberg and Weiss. *See* Ginsberg & Weiss, *supra* note 3, at 932.

5. The 301(e) Study recommended the creation of rebuttable presumptions of causation in favor of plaintiffs seeking compensation from an administrative fund. See *supra* note 3. While the Study did not recommend that its presumptions carry over to tort actions, neither did it recommend against such a step. 301(e) Study, *supra* note 3, at 260. Some commentators have affirmatively proposed this. *See, e.g.*, Burcat, *Uncompensated Victims of Low-Level Radiation: Unnecessary Hostages of the Price-Anderson Act Debate*, 15 Forum 847, 859 (1980); Delgado, *Beyond* Sindell: *Relaxation of Cause-In-Fact Rules for Indeterminate Plaintiffs*, 70 Calif. L. Rev. 881, 899 (1982); Note, *The Inapplicability of Traditional Tort Analysis to Environmental Risks: The Example of Toxic Waste Pollution Victim Compensation*, 35 Stan. L. Rev. 575, 615 (1983) [hereinafter cited as *Environmental Risks*]; Note, *Tort Actions for Cancer; Deterrence, Compensation, and Environmental Carcinogenesis*, 90 Yale L.J. 840, 855 (1981) [hereinafter cited as *Tort Actions for Cancer*]. Other commentators have suggested evidentiary standards that are stacked in favor of plaintiffs though not couched in terms of presumptions. *See, e.g.*, Hall & Silbergeld, *Reappraising Epidemiology: A Response to Mr. Dore*, 7 Harv. Envtl. L. Rev. 441, 444-45 (1983); *Precursor Symptoms*, *supra* note 2, at 189-90.

Some commentators have proposed proportional liability as an alternative to the traditional all-or-nothing recovery approach. *See, e.g.*, Delgado, *supra*, at 899-902; Rizzo & Arnold, *Causal Apportionment in the Law of Torts; An Economic Theory*, 80 Colum. L. Rev. 1399, 1407-13 (1980); Robinson, *Multiple Causation in Tort Law: Reflections on the DES Cases*, 68 Va. L. Rev. 713, 755-58 (1982); Rosenberg, *The Causal Connection in Mass Exposure Cases: A "Public Law" Vision of the Tort System*, 97 Harv. L. Rev. 849, 881-87 (1984). The theory underlying these proposals

This Article focuses on the use of the traditional preponderance-of-the-evidence standard of proof in toxic tort cases in which a single substance is at issue.[6] Courts have found it difficult to apply this standard to the kind of evidence seen in toxic tort litigation, and as a result, have sometimes allowed recovery based on highly suspect evidence,[7] or conversely, have failed adequately to justify the exclusion of evidence.[8]

---

has not been developed fully, however, and such a reform may be inappropriate for single-factor cases. Multi-factor cases may provide stronger justification for proportional liability, but its rational use would require the application of evidentiary tests very similar to the one proposed in this Article. See *infra* pt. V(B).

6. While many cases involve allegations that a number of substances combined to cause a plaintiff's disease, the analysis in this Article is confined to cases in which a plaintiff is exposed to a single identifiable substance and subsequently contracts a disease. The disease is known to arise without the identified exposure, but the plaintiff nonetheless links his or her case to that exposure. Examples include litigation about asbestos, Agent Orange and radiation. Of course, all of these cases involve at least two factors: the substance at issue and whatever other factor(s) (for example, diet or exposure to other substances) are responsible for the background incidence rate of the disease. They are single-factor cases in the legal sense, however, because liability will attach, if at all, to the one identifiable factor.

Cases in which the accused substance allegedly interacted with other factors involve issues not addressed in this Article. Also unaddressed are cases involving two or more identifiable factors, each independently sufficient to cause the injury at issue. When more than one factor is a source of potential liability, however, the epidemiologic concept of attributable risk, upon which this Article is based, still provides the only scientifically valid factual basis for legal analysis. For example, a lung cancer victim exposed to both benzene and cigarettes might be able to attribute 60% of the risk to cigarettes, 20% to benzene, and 20% to unknown factors. These attributable risks are either additive or multiplicative (in lay terms "synergistic"). If the former, the analysis of this Article can be applied with little further elaboration; if the latter, other rules of attribution are required.

7. The full extent of this problem is not revealed by published decisions. Often when a plaintiff with an unsubstantiated claim wins a verdict after presenting very questionable evidence, the defendant will simply settle. The lack of clear standards turns appellate review into a crapshoot with the dice loaded for the plaintiff. A recent example is instructive. In Grasso v. B.F. Goodrich Co., No. 78-1562 (D.N.J. Jan. 30, 1981), the plaintiff alleged that his liver cancer (angiosarcoma) had been caused by vinyl chloride (VC) from a factory located near his home. The expert witness called by the plaintiff to establish this theory of "neighborhood cancer" testified that, in addition to the plaintiff, eight documented cases of angiosarcoma had occurred within two miles of an industrial plant using VC. Trial Transcript at 97, *Grasso*. The expert, however, did not substantiate his conclusion about causation. Although he acknowledged that angiosarcoma could occur without any exposure, *id.* at 182, and that 75% of all cases were of unknown origin, *id.* at 121, he took the position that diagnosis of angiosarcoma and proximity to an emitting VC source would suffice to establish VC as the cause. *Id.* at 177-78. He also seemed unwilling or unable to distinguish between an explanation that is "more likely than not" correct and one that is "the most likely" of several explanations. *Id.* at 112-13. Despite this weak evidence, the jury returned a plaintiff's verdict that the trial court refused to set aside. An appeal was taken, but the case was settled before argument. A clear

These problems can be overcome, however, if courts apply recognized epidemiologic principles and concepts in conjunction with the traditional standard of proof. Epidemiology is the only generally accepted scientific discipline that deals with the integrated use of statistics and biological/medical science to identify and establish the causes of human diseases.[9] Its use enables scientific estimation of the percentage of the risk of a disease that is properly attributable to a given factor, such as exposure to an allegedly harmful substance. Thus, use of an epidemiologic standard would provide courts with a rational and consistent means for evaluating evidence of a causal relationship between exposure to a particular factor and the incidence of a disease.

---

evidentiary standard might have prevented the initiation of a case like *Grasso*, and surely would have made it easier to take the case from the jury or to argue for reversal on appeal.

Counsel to one major chemical company has publicly lamented the ease with which plaintiffs can obtain settlements in toxic exposure cases. He attributes the problem to complexity and expense of defense as well as to the uncertainty of the outcome at trial. Sheridan, *Rethinking Mass Tort Defense*, Litigation, Summer 1983, at 29, 29-30.

8. Two recent cases in the District of Columbia, both involving the anti-morning sickness drug Bendectin, illustrate the problems courts have in justifying the exclusion of patently insufficient evidence at the outset of a trial, or in taking a case from the jury if no other evidence is introduced during the trial. In Koller v. Richardson-Merrell, Inc., No. 80-1258 (D.D.C. filed Feb. 25, 1983) the plaintiff alleged that her birth defects had been caused by Bendectin manufactured by the defendant and taken by her mother during pregnancy. The court, in a preliminary order, required that all statistical evidence be significant at a 95% confidence level. *Id.* at 1. This kept certain causation testimony out of the trial, which the plaintiff lost. Neither the order nor the memorandum opinion indicate what is meant by "significance," however. If the reference is simply to the existence of a significant difference between children of mothers who took Bendectin and those who did not, the ruling makes sense, but if the reference goes to complicated statistics such as the risk ratio, significance testing makes little sense. See *infra* pt. II.

In Oxendine v. Merrell Dow Pharmaceuticals, Inc., No. 1245-82 (Super. Ct. D.C. filed Sept. 1, 1983), the judge allowed the testimony that had been excluded in *Koller*. After a jury verdict for the plaintiff, however, the court granted a judgment n.o.v. The judge found no evidence that Bendectin could cause birth defects, *id.* at 2, although the plaintiff's expert testified that 21 of 1,000 children born to mothers who had taken Bendectin would have defects compared to no more than 20 of 1,000 children born to mothers who had not. Trial Transcript at 108-09, *Oxendine*. These statistics are evidence that the drug causes some birth defects, though at most only a very small percentage. The judge would have been on much firmer ground had he found the evidence insufficient to satisfy the more-likely-than-not test rather than finding that it showed nothing at all.

9. Epidemiology is a well-established science tracing its roots back at least 150 years. While not a required part of the typical medical school curriculum, it is taught at most schools. Epidemiologists are not necessarily medical doctors, but many do have M.D.'s. The discussion of epidemiology, *infra* pt. II, explains at some length the discipline's relationship to other sciences.

This Article's underlying premise is that a toxic tort plaintiff, like any other tort plaintiff, has the burden of proving each element of his case[10], including causation.[11] This burden includes the production of evidence from which the factfinder could reasonably infer that the accused substance "more likely than not" caused the plaintiff's harm.[12] The plaintiff must introduce evidence of both the substance's harmfulness at a given exposure level, and of his exposure to the

---

10. "Burden of proof" is an unfortunately ambiguous term that incorporates both the burden of producing evidence and the ultimate burden of persuasion. *See* Laughlin, *The Location of the Burden of Persuasion*, 18 U. Pitt. L. Rev. 3, 3 (1956). Inasmuch as this Article deals with the sufficiency of evidence, it is about the burden of production. *See* Dworkin, *Easy Cases, Bad Law, and Burdens of Proof*, 25 Vand. L. Rev. 1151, 1160 (1972). Courts and commentators have considered a number of factors in discussions of how the burden of proof (production or persuasion) should be allocated. These can be grouped under a few broad headings: probability, access to evidence and policy. See *infra* notes 135-42 and accompanying text. The general rule is that the "burdens of pleading and proof with regard to most facts have been and should be assigned to the plaintiff who generally seeks to change the present state of affairs and who therefore naturally should be expected to bear the risk of failure of proof or persuasion." E. Cleary, McCormick's Handbook on the Law of Evidence § 337, at 786 (2d ed. 1972).

11. Like "burden of proof," "causation" has been the source of much confusion. The law distinguishes between "cause in fact" and "proximate cause." The former is simply a matter of what has, in fact, occurred. *See* W. Prosser, Law of Torts § 41, at 237 (4th ed. 1971). The latter is a matter of law. *Id.* § 42, at 244. This Article is concerned solely with the issue of cause in fact, on which:

> as on other issues essential to his cause of action for negligence, the plaintiff . . . has the burden of proof. He must introduce evidence which affords a reasonable basis for the conclusion that it is more likely than not that the conduct of the defendant was a substantial factor in bringing about the result.

Id. § 41, at 241 (footnote omitted). Although this language seems restricted to negligence actions, Dean Prosser made clear that causation is also an essential element for any other tort. *Id.*

The "substantial factor" concept was developed to enable the law to deal with situations in which two or more factors combine to bring about a plaintiff's injury. It does not apply to cases in which factors have acted independently. *See* Delgado, *supra* note 5, at 886-87 & n.26 (referring to "material and contributing" factors, but citing the discussion of "substantial factors" in W. Prosser, *supra*, § 41, at 240-41). Because this Article is restricted to fact patterns involving a single identifiable factor, the "substantial factor" element in Dean Prosser's analysis need not be addressed.

12. Even commentators who have advocated changes to make it easier for plaintiffs to recover in toxic tort cases have explicitly recognized that the more-likely-than-not test is the present rule. *See* Hall & Silbergeld, *supra* note 5, at 446; Trauberman, *supra* note 5, at 197; *Environmental Risks*, *supra* note 5, at 578; *Tort Actions for Cancer*, *supra* note 5, at 857 n.77; *see also Precursor Symptoms*, *supra* note 2, at 193, in which the author explicitly states that the Note's theory requires placing the burden of uncertainty on defendants. The rationale for placing the burden of proof on plaintiffs and for requiring evidence sufficient to establish that the plaintiffs' allegations are more likely than not true is discussed *infra* pt. III(A).

substance at or above that level.[13] Because most toxic tort cases involve diseases with long latency or incubation periods, and because many of these diseases may occur in the absence of any identifiable exposure, causation very often becomes a central and complex issue at trial.[14] To resolve this issue, plaintiffs usually must resort to expert witnesses[15] who, unfortunately, sometimes venture opinions unsupported by scientific data.[16] Moreover, while the outcome of many cases depends on the legal sufficiency of such evidence,[17] courts have not been able to decide the sufficiency issue either clearly or consistently.

---

13. *See* 301(e) Study, *supra* note 3, at 70-71. The standard this Article proposes pertains principally, but not exclusively, to the harmfulness aspect of causation. Unlike proof of harmfulness, proof of individual exposure generally depends on more traditional evidence. For a case that turned on the distinction between harmfulness and exposure, see Besner v. Walter Kidde Nuclear Lab., 18 A.D.2d 952, 952, 237 N.Y.S.2d 585, 587 (1963) (holding that the plaintiff had not established causation because the only expert witness who testified about a causal relationship "based his opinion on a completely erroneous premise as to the length of exposure involved and/or a set of facts as to the amount, nature or duration of the alleged exposure unsubstantiated by the record"). The case was remanded and the plaintiff won again below. The defendant once more appealed, but the plaintiff prevailed. He had been able to establish exposure "for a substantial part of two periods and also at other times in various amounts." Besner v. Walter Kidde Nuclear Lab., 24 A.D.2d 1045, 1045, 265 N.Y.S.2d 312, 313 (1965).

14. *See Tort Actions for Cancer*, *supra* note 5, at 851-55; *see, e.g.*, Boldt v. Josten's, Inc., 261 N.W.2d 92, 94 (Minn. 1977); Miller v. National Cabinet Co., 8 N.Y.2d 277, 282-83, 168 N.E.2d 811, 813-14, 204 N.Y.S.2d 129, 132-33 (1960); Clark v. Workmen's Comp. Comm'r, 155 W. Va. 726, 731-34, 187 S.E.2d 213, 216-18 (1972).

15. *See* Taylor, *Occupational Disease: A Defense Attorney's Point of View*, 12 Forum 297, 299 (1976); Trauberman, *supra* note 3, at 189 n.4.

16. An example of this is provided by an expert who testified in several of the swine flu cases. See *infra* pt. IV(B)(3). In one case he opined that the plaintiff's arthritis had been caused by her swine flu inoculation. Gicas v. United States, 508 F. Supp. 217, 220 (E.D. Wis. 1981). The court found:

> that the overwhelming weight of the medical literature opposes a theory that associates Swine Flu vaccine to the plaintiff's injuries. No authority other than [the expert] has causally related rheumatoid arthritis with a swine flu inoculation. . . . [The expert] knows of no evidence other than this case that supports his theory.

Id. Faced with the same expert's testimony in another case, the court noted that "[t]he posture of the expert testimony in this case indicates the limited usefulness that such testimony offers a trier of fact." Latinovich v. United States, 537 F. Supp. 671, 676 (E.D. Wis. 1982). The court went on to list a number of other cases in which his theories had been rejected. *Id.* This expert was also explicitly rejected in Kubs v. United States, 537 F. Supp. 560, 563 (E.D. Wis. 1982).

17. A distinction must be drawn between sufficiency and admissibility. Insufficient evidence may be admissible, but if this is all that a plaintiff can offer, as a matter of law, he cannot prevail. For discussions of the distinction between sufficiency and admissibility, see Martin, *The Uncertain Rule of Certainty: An Analysis*

The first part of this Article examines the inconsistencies and deficiencies in cases that have addressed the issue of causation. Courts have recognized the need to infer causation in toxic tort cases from differences between exposed and unexposed populations. At the same time, they have tried to hold to basic tort law principles. Without a test to measure causal inferences against legal principles, however, their decisions have been ambiguous and confusing. The second part of the Article provides an introduction to the principles of epidemiology, which form the basis for a proposed standard that will enable courts better to distinguish insufficient from sufficient evidence. Part III establishes the basis for the premise that the preponderance-of-the-evidence standard should apply in toxic tort cases, and it then combines epidemiologic principles with this premise to formulate a standard for determining evidentiary sufficiency in toxic tort cases. The proposed standard would require the plaintiff to establish that more than fifty percent of the risk of developing the disease at issue be attributable to the substance at issue, and that certain fundamental epidemiologic postulates be satisfied. Part IV discusses precedents for the use of epidemiologic principles by courts, and possible requirements for witnesses who testify as expert epidemiologists. Finally, the Article addresses problems that might result from retaining the traditional burden of proof and using an evidentiary standard that requires the accumulation of data about populations before an individual can bring a successful action.

I. Causation in Cancer and Toxic Tort Cases

A. *Cancer Cases Involving Trauma or Irritation*

Legal inquiry into the causation of cancer pre-dates toxic tort law, and much of the early theory persists today. Plaintiffs often allege causation from either a traumatic injury[18] or exposure to an immedi-

---

and Proposal For a Federal Evidence Rule, 20 Wayne L. Rev. 781, 797-802 (1974); Musslewhite, *Medical Causation Testimony in Texas: Possibility vs. Probability*, 23 Sw. L.J. 622, 622 (1969); Note, *Causation in Disease: Quantum of Proof Required to Reach the Jury*, 53 Nw. U.L. Rev. 794, 795-98 (1959).

18. *E.g.*, Kramer Servs., Inc., v. Wilkins, 184 Miss. 483, 496, 186 So. 625, 627 (1939) (plaintiff alleged that his cancer had been caused by a cut he received when broken glass fell on him); Stordahl v. Rush Implement Co., 148 Mont. 13, 14-16, 417 P.2d 95, 96-97 (1966) (cancer allegedly caused by blow to back); Casson v. A.C. Horn Co., 27 A.D.2d 966, 966-67, 279 N.Y.S.2d 244, 245 (1967) (lung cancer allegedly caused by inhaling paint fumes in work place accident); Hanna v. Aetna Ins. Co., 24 Ohio Misc. 27, 28, 259 N.E.2d 177, 178 (1970) (breast cancer allegedly caused by bruises suffered in automobile accident); Gambrell v. Burleson, 252 S.C. 98, 100, 165 S.E.2d 622, 622-23 (1969) (cancer allegedly aggravated by automobile accident). Most of the injuries are single, isolated traumas, though some are repeated

ately irritating or harmful substance, such as sand or sulfuric acid.[19] In adjudicating trauma claims, courts usually fail to recognize that cancers generally develop without identifiable prior traumatic events, and that incidence rates are no higher in groups that have suffered single traumatic injuries than in those that have not.[20] While appellate decisions sometimes acknowledge the uncertainty and ignorance that surround cancer, they often uphold plaintiffs' verdicts based on coincidences lacking statistical significance.[21] What little guidance medical science has provided about traumatic causation is frequently ignored or misinterpreted.

In 1926, Dr. James Ewing outlined criteria for attributing a particular cancer to a trauma.[22] Although these criteria were intended to provide guidance to courts, Ewing cautioned that "[t]he traumatic theory runs against too many general objections to permit its uncriti-

---

traumas more akin to physical irritation. For purposes of this Article, trauma will mean single trauma.

19. *E.g.*, Hagy v. Allied Chem. & Dye Corp., 122 Cal. App. 2d 361, 363, 265 P.2d 86, 87 (1953) (cancer allegedly caused or aggravated by exposure to sulfuric acid); Bollinger v. Wagaraw Bldg. Supply Co., 122 N.J.L. 512, 514-15, 6 A.2d 396, 398-99 (1939) (plaintiff claimed that sand and ashes that had gotten into the decedent's shoes had so aggravated a pigmented mole on one of his feet that it developed into a cancer); Chalmers v. Dep't of Labor & Indus., 72 Wash. 2d 595, 597, 434 P.2d 720, 721 (1967) (cancer allegedly caused by fumes so irritating they once caused plaintiff's deceased husband to pass out); *see* Adelson, *Injury and Cancer*, 5 W. Res. L. Rev. 150, 168-69 (1954); Dyke, *Traumatic Cancer*, 15 Clev.-Mar. L. Rev. 472, 484-94 (1966); Comment, *Sufficiency of Proof in Traumatic Cancer: A Medico-Legal Quandary*, 16 Ark. L. Rev. 243, 256-67 (1962); Comment, *Judicial Attitudes Towards Legal and Scientific Proof of Cancer Causation*, 3 Colum. J. Envtl. L. 344, 354-68 (1977) [hereinafter cited as *Scientific Proof*]; Comment, *Sufficiency of Proof in Traumatic Cancer Cases*, 46 Cornell L.Q. 581, 581-82 (1961) [hereinafter cited as *Sufficiency of Proof*].

20. *See* Adelson, *supra* note 19, at 154-55; Auster, *The Role of Trauma in Oncogenesis: A Juridical Consideration*, 175 J. A.M.A. 946, 949 (1961); Russell & Clark, *Medico-Legal Considerations of Trauma and Other External Influences in Relationship to Cancer*, 6 Vand. L. Rev. 868, 875 (1953); Warren, *Criteria Required to Prove Causation of Occupational or Traumatic Tumors*, 10 U. Chi. L. Rev. 313, 318-20 (1943).

21. *E.g.*, Hagy v. Allied Chem. & Dye Corp., 122 Cal. App. 2d 361, 375-76, 265 P.2d 86, 95 (1953); Daly v. Bergstedt, 267 Minn. 244, 248, 126 N.W.2d 242, 245 (1964); *see Sufficiency of Proof*, *supra* note 19, at 582 & n.10. See *infra* note 104 for a discussion of what is meant by "statistical significance."

22. Ewing, *The Relation of Trauma to Malignant Tumors*, Am. J. Surgery, Feb. 1926, at 30, 31-34. The criteria set forth were:

(1) Authenticity and sufficient severity of the trauma.

(2) Previous integrity of wounded part.

(3) Identity of injured area with that giving origin to the tumor.

(4) Tumor of a type that could conceivably result from trauma.

(5) Proper time interval between receipt of the injury and appearance of the tumor.

Id.

cal acceptance."[23] Moreover, he premised his work on the assumption that the defendant has the burden of disproof,[24] thus further limiting the proper application of his postulates. By 1935, he had become still more conservative, acknowledging that "experimental data reveal the fact that cancer genesis requires quite peculiar factors which have not been found in the results of simple trauma."[25] Later work by others has further limited the Ewing approach.[26]

Ignorance and uncertainty make it virtually impossible, even with the aid of Ewing's criteria, to determine whether a single trauma, or a majority of irritating factors,[27] more likely than not caused the initiation of a latent disease such as cancer. Because plaintiffs bear the burden of proof, this dearth of evidence logically implies that plaintiffs should generally lose as a matter of law, but few courts have stated this explicitly.[28] Instead, decisions have generally been ill-reasoned and inconsistent.[29]

---

23. *Id.* at 34.

24. *Id.* at 30.

25. Ewing, *The Modern Attitude Toward Traumatic Cancer*, 11 Bull. N.Y. Acad. Med. 281, 281 (1935).

26. *See* Auster, *supra* note 20, at 949. No one has suggested that the Ewing analysis can lead to a conclusion that a causal link is more probable than not. Rather, only possible inference is claimed. One commentator has explicitly stated that the postulates relate only to possibility. Adelson, *supra* note 19, at 156. Ewing's postulates may, however, be used to support defendants' verdicts because the plaintiff must at least satisfy them to prove causation. *See* Stordahl v. Rush Implement Co., 148 Mont. 13, 19-20, 417 P.2d 95, 99 (1966); Sikora v. Apex Beverage Corp., 282 A.D. 193, 196, 122 N.Y.S.2d 64, 66 (1953), *aff'd*, 306 N.Y. 913, 119 N.E.2d 601 (1954); Dennison v. Wing, 279 A.D. 494, 496-97, 110 N.Y.S.2d 811, 813-14 (1952).

27. *See* Auster, *supra* note 20, at 949. In some prolonged irritation cases it may be possible to infer causation with sufficient certainty. Ewing, *supra* note 25, at 314.

28. The only example of which the authors are aware is Tonkovich v. Department of Labor & Indus., 31 Wash. 2d 220, 195 P.2d 638 (1948).

29. *Compare* Daly v. Bergstedt, 267 Minn. 244, 248, 126 N.W.2d 242, 245 (1964) (upholding plaintiff's claim that a bruise on her breast had become cancerous) *with* Tonkovich v. Department of Labor & Indus., 31 Wash. 2d 220, 226-27, 195 P.2d 638, 641-42 (1948) (rejecting plaintiff's claim that fractured bones in his foot worsened into arthritis and intestinal cancer 10 years later).

Plaintiffs' verdicts in workers' compensation cases, even in the absence of reliable information, are perhaps understandable. The requirement that a disease be occupational conceptually parallels the tort law causation requirement, but it is not identical to it. *See* 1B A. Larson, Workmen's Compensation Law § 41 (1982 & Supp. 1983); *see, e.g.*, Cox v. Ulysses Coop. Oil & Supply Co., 218 Kan. 428, 432-33, 544 P.2d 363, 367 (1975) (in a workers' compensation case the claimant need only introduce evidence sufficient to convince the court that the award is proper); Deines v. Greer, 216 Kan. 548, 553, 532 P.2d 1257, 1262 (1975) (when injury shown to have arisen out of course of employment, every natural consequence of injury is compensable); Workmen's Comp. Appeals Bd. v. Bethlehem Steel Corp., 23 Pa. Commw. 454, 456, 352 A.2d 571, 572 (1976) (plaintiff need not prove injury caused by identifiable incident, but rather only that injury arose in course of employment). Some states

Daly v. Bergstedt[30] typifies the muddled reasoning employed in many trauma and irritation cases. The plaintiff brought a simple slip and fall tort action, straightforward except for her claim that a bruise on her left breast had caused it to become cancerous. The Minnesota Supreme Court affirmed the plaintiff's verdict, but the court's review of the evidence did not justify its holding. Six medical doctors testified that there was no causal connection between the bruise and the cancer, while one gave the opinion that the cancer could have developed from the trauma sustained in the fall.[31] Apparently realizing that science weighed heavily in favor of the defendant, the court chose to rely on the coincident location of the trauma and the cancer and the relatively short (14 months) time period between the two.[32] This approach totally ignores the absence of evidence that the incidence of breast cancer is higher among women who have suffered trauma than among women who have not.[33] The *Daly* case implies that it is appropriate to allow laymen to draw conclusions from information

---

create presumptions that lessen the plaintiff's burden of proof. *See, e.g.*, Downes v. Industrial Comm'n, 113 Ariz. 90, 93, 546 P.2d 826, 829-30 (1976); Bolger v. Chris Anderson Roofing Co., 112 N.J. Super. 383, 394, 271 A.2d 451, 457-58 (1970), *aff'd per curiam*, 117 N.J. Super. 497, 285 A.2d 228 (1971). *Compare* Cox v. Ulysses Coop. Oil & Supply Co., 218 Kan. 428, 435-36, 544 P.2d 363, 369-70 (1975) (personal opinion of physician that causation is a "reasonable medical certainty" is sufficient to justify recovery) *with* Parker v. Employers Mut. Liab. Ins. Co., 440 S.W.2d 43, 45 (Tex. 1969) (causal connection must be clearly established between employment and injury to justify recovery).

30. 267 Minn. 244, 126 N.W.2d 242 (1964).

31. *Id.* at 248, 126 N.W.2d at 245. The court based its opinion on the Ewing Postulates. *Id.* However, the postulates had not, in fact, been satisfied. Ewing made it quite clear that only one type of breast cancer, carcinoma simplex, could be linked to trauma, and that "in each case the entire clinical history must be secured and the tumor and the entire breast must be examined by a competent tumor pathologist before the basis can be laid for an opinion." Ewing, *supra* note 25, at 320-21. There is no indication that Mrs. Daly produced such evidence. In fact, her expert had testified that she had a scirrhus carcinoma, not carcinoma simplex. 267 Minn. at 249, 126 N.W.2d at 246. Even if the postulates had been satisfied, the plaintiff would still not have established the causal link by a preponderance of the evidence. See *supra* note 10.

32. 267 Minn. at 247-51, 126 N.W.2d at 245-47. Other courts have held that while coincidence and expert testimony about possibilities by themselves are not enough, together they may be sufficient. *See* Hagy v. Allied Chem. & Dye Corp., 122 Cal. App. 2d 361, 371, 265 P.2d 86, 92-93 (1953) (quoting Fireman's Fund Indemnity Co. v. Industrial Acc. Comm'n, 93 Cal. App. 2d 244, 246, 208 P.2d 1033, 1034 (1949)). While plausible at first glance, this approach is in fact no better than that taken by the *Daly* court. An expert is assumed to know all the available facts relevant to causation, and if he cannot reach a suitably certain conclusion laymen should not be expected to do so. Stated another way, if proof of causation requires expert testimony, the expert's determination of how certain one can be ought to be determinative.

33. See *supra* note 20 and accompanying text.

found to be inadequate by experts, a rule that leaves little basis for a rational analysis of the legal sufficiency of evidence.[34]

Courts that have reviewed the sufficiency of expert testimony in trauma and irritant cases have tended to go little beyond the witness' expressed degree of certainty, distinguishing, for example, between the use of the words "possible" and "probable."[35] Often they uncritically defer to physicians,[36] whose training and experience typically do not qualify them to venture opinions about the probability that a particular factor caused a disease.[37] Focusing on the expressed certainty or supposed professional competence of physicians shifts attention from underlying uncertainty and permits at least apparent adherence to the more-likely-than-not standard, but it does not lead to consistent results.

The distinction between possibility and probability is not insignificant, but when reduced to a simple search for expressed certainty or for the blessing of a suitably credentialed expert, it often has no real effect. Judicial reluctance to examine the substantive basis of the testimony can easily permit unfounded expressions of certainty to carry the day. Pennsylvania, for example, requires that causation

---

34. Other decisions have also been based on this kind of limited review. *See, e.g.*, Hagy v. Allied Chem. & Dye Corp., 122 Cal. App. 2d 361, 375-76, 265 P.2d 86, 95 (1953); Hanna v. Aetna Ins. Co., 24 Ohio Misc. 27, 32-33, 259 N.E.2d 177, 180-81 (1970); Valente v. Bourne Mills, 77 R.I. 274, 278, 75 A.2d 191, 194 (1950).

35. Cox v. Ulysses Coop. Oil & Supply Co., 218 Kan. 428, 435-36, 544 P.2d 363, 369-70 (1975); *see* Pucci v. Rausch, 51 Wis. 2d 513, 518-19, 187 N.W.2d 138, 141-42 (1971) (personal injury case in which the court required only that a doctor have sufficient certainty that his opinion is "correct to a reasonable medical probability. Other doctors may differ, but whether his opinion corresponds with that of another member of the medical profession does not go to admissibility of his opinion but to the weight the trier of the facts should give to his opinion."); City of Seymour v. Industrial Comm'n, 25 Wis. 2d 482, 491-92, 131 N.W.2d 323, 328 (1964) (medical testimony cannot be held "incredible because contrary to scientific facts or knowledge").

The *Pucci* court listed with approval a number of cases in which various forms of medical testimony had been either acceptable or unacceptable. 51 Wis. 2d at 519, 187 N.W.2d at 142. This approach can redound to the benefit of defendants as well as plaintiffs. *See* Casson v. A.C. Horn Co., 27 A.D.2d 966, 967, 279 N.Y.S.2d 244, 245 (1967) (medical testimony sufficient); Insurance Co. of N. Am. v. Myers, 411 S.W.2d 710, 714 (Tex. 1966) (medical testimony that causation was merely possible insufficient for recovery). *See generally* Annot., 66 A.L.R.2d 1082, 1118-24 (1959) (dealing with the issue of admissibility, not sufficiency, but citing many cases that relate to the sufficiency issue).

36. *See* McGrath v. Irving, 24 A.D.2d 236, 238, 265 N.Y.S.2d 376, 378 (1965) (plaintiff's expert testimony held sufficient based on his "medical qualifications").

37. When etiology is unknown, causation must usually be determined at least in part from statistical inferences. Biostatisticians deal with this numerical aspect of establishing causation, but they often lack a full appreciation of the biological aspect. It is the epidemiologist who specializes in using both statistics and biology to arrive at scientifically supportable conclusions about causation.

testimony be couched in very certain terms, but an expert in *Menarde v. Philadelphia Transportation Co.*[38] evaded this limitation simply by testifying that it was virtually impossible that the plaintiff's breast cancer had been caused by anything other than the minor injuries she had suffered in a trolley car accident.[39] This case clearly demonstrates how neatly an expert can tailor testimony to the requirements set forth in previous decisions. If certainty is needed, witnesses can be found who will profess it.

B. *Toxic Tort Cases*

In toxic tort cases, latency and the absence of an identifiable irritation or traumatic injury have made it more difficult than in trauma cases for courts to rely solely on coincidences.[40] Nevertheless, the focus on witnesses' expressions of certainty and the deference to medical experts seen in traumatic cancer cases have carried over to toxic torts. In *Boldt v. Jostens, Inc.*,[41] for example, the plaintiff claimed that her workplace exposure to fumes from heated glue caused her to contract Goodpasture's Syndrome, a pathologic condition in which the kidneys and lungs are attacked by one's own immune system. The doctor who testified for the plaintiff about causation acknowledged that the etiology of Goodpasture's Syndrome is unknown. He stated that it was thought to be an immunologic disease, that the antigen causing a reaction in a victim "can probably be many different things and different for different people,"[42] and that it is unknown whether the reaction is the result of one exposure or many.[43] Yet, he was willing to opine that the plaintiff's exposure to glue fumes "had a great deal to do with her illness, and certainly caused aggravation."[44] The Supreme Court of Minnesota held that this testimony sufficed to sustain a workers' compensation award, in part because "the truth of the opin-

---

38. 376 Pa. 497, 103 A.2d 681 (1954).

39. *Id.* at 502, 103 A.2d at 684; *see* Peterson v. Kansas City Pub. Serv. Co., 259 S.W.2d 789, 794 (Mo. 1953).

40. *Scientific Proof*, *supra* note 19, at 354. *But see* Boney v. Gouverneur Talc Co., 77 A.D.2d 702, 702, 430 N.Y.S.2d 399, 399 (1980) (lung cancer found to have been caused by exposure to talc dust containing asbestos). The plaintiff in *Boney* admittedly had talcosis, a form of pneumoconiosis. But other than testimony that this condition might have predisposed him to contract cancer, there was apparently no evidence to support holding the defendent liable for the disease. It should be noted that while mesothelioma (which is not what the plaintiff had) is very clearly linked to asbestos exposure, other forms of cancer are not, at least not to the same high degree. This illustrates why specificity is so important in epidemiologic analysis. See *supra* note 89 and accompanying text.

41. 261 N.W.2d 92 (Minn. 1977).

42. *Id.* at 93.

43. *Id.*

44. *Id.*

ion need not be capable of demonstration."[45] Other cases indicate that in some circumstances a treating physician's testimony will be given special weight,[46] or that a specialist's testimony will be given more weight than a general practitioner's.[47]

When courts do go beyond simple deference to medical testimony, they generally do no more than subject it to the same cursory "probability versus possibility" analysis found in some trauma cases.[48] In exposure cases this has at least proven useful in culling claims in which a witness singles out one factor as the "most probable" of many. These situations occur because when diagnosing and treating a disease, doctors often cannot state with certainty which factor is its direct cause. They quite properly think in terms of finding the most likely cause instead of a factor that more likely than not is the cause.[49] Thus terms like "medical certainty" or "medical probability" often fail to satisfy legal requirements. In *Clark v. State Workmen's Compensation Commissioner*,[50] for example, the plaintiff established that the only clearly identifiable cause of her deceased husband's leukemia was his exposure to chemicals at the plant in which he had worked.[51] Her expert had also testified, however, that the etiology of the disease was unknown and that other factors could have caused it.[52] The court held that this evidence failed to satisfy the requirement that a workers' compensation claimant prove that his disease is job-related.[53]

Although scientific studies do not support the argument that trauma increases the incidence of disease,[54] data do exist that permit comparisons of disease rates in populations exposed to some substances with the rates in unexposed populations.[55] Such comparisons are a

---

45. *Id.* at 94. *But see* Logan Co. v. Amic, 479 S.W.2d 1, 2-3 (Ky. 1972) (hypothesis of physician not sufficent evidence to justify plaintiff's recovery).

46. Long v. Martin Timber Co., 395 So. 2d 931, 934 (La. App. 1981); Groff v. Department of Labor & Indus., 65 Wash. 2d 35, 45, 395 P.2d 633, 639 (1964); *Sufficiency of Proof, supra* note 19, at 601.

47. Chalmers v. Department of Labor & Indus., 72 Wash. 2d 595, 598-601, 434 P.2d 720, 722-24 (1967); *Sufficiency of Proof, supra* note 19, at 601.

48. See *supra* notes 35-39 and accompanying text.

49. *See* Danner & Sagall, *Medicolegal Causation: A Source of Professional Misunderstanding*, 3 Am. J. Law & Med. 303, 304-05 (1977).

50. 155 W. Va. 726, 187 S.E.2d 213 (1972).

51. *Id.* at 728-29, 187 S.E.2d at 215.

52. *Id.*

53. *Id.* at 734, 187 S.E.2d at 217-18; *see* Schaefer v. Texas Employment Ins. Ass'n, 612 S.W.2d 199, 205 (Tex. 1981) (rejecting medical testimony that it was reasonably probable that workplace exposure caused decedent's cancer).

54. See *supra* note 20 and accompanying text.

55. *See, e.g.*, Doll & Peto, *The Causes of Cancer: Quantitative Estimates of Avoidable Risks of Cancer in the United States Today*, 66 J. Nat'l Cancer Inst. 1192

basic part of epidemiologic studies, and in a few cases, have led to causal inferences so strong that courts have found causation to be scientifically established without any analysis of the method and reasoning underlying that conclusion.[56] When the data are less conclusive, as usually occurs in toxic tort cases, the law has had far more difficulty in dealing with the evidence. A number of commentators have referred approvingly to the use of epidemiology or biostatistics,[57]

(1981); Wynder & Gori, *Contribution of the Environment to Cancer Incidence: An Epidemiologic Exercise*, 57 J. Nat'l Cancer Inst. 825 (1977).

56. The link between asbestos and mesothelioma (a form of cancer that attacks the lining of the pleural cavity) was not established until the early 1970's, just about the time that the growing flood of legal action began. Mehaffy, *Asbestos-Related Lung Disease*, 16 Forum 341, 344 (1980). Causation had been established by the epidemiologic work of Dr. Irving J. Selikoff and others, and in the litigation it has been substantially accepted. Without examining the methodology by which scientists reached their conclusions, courts accept causation almost as a matter of law. *See* Karjala v. Johns-Manville Prods. Corp., 523 F.2d 155, 158 (8th Cir. 1975); Bertrand v. Johns-Manville Sales Corp., 529 F. Supp. 539, 544 (D. Minn. 1982); Flatt v. Johns-Manville Sales Corp., 488 F. Supp. 836, 841 (E.D. Tex. 1980); Mehaffy, *supra*, at 341. *But see* Tretter v. Johns-Manville Corp., 88 F.R.D. 329, 332-33 (E.D. Mo. 1980) (court required plaintiff asserting causal link between asbestos and cancer to prove harmfulness).

In the DES litigation, the link between DES and clear cell adenocarcinoma is virtually certain, although established only epidemiologically. *See* Herbst, Ulfelder & Poskanzer, *Adenocarcinoma of the Vagina*, 284 N.E. J. Med. 878, 878 (1971); Note, *Market Share Liability: An Answer to the DES Causation Problem*, 94 Harv. L. Rev. 668, 669 (1981). Vinyl chloride exposure (at high enough levels) and one form of liver cancer have also been linked almost unequivocally through epidemiology. *See* Society of the Plastics Indus., Inc. v. OSHA, 509 F.2d 1301, 1305-06 (2d Cir.), *cert. denied*, 421 U.S. 992 (1975).

In the case of cigarettes and lung cancer, some early decisions indicated that epidemiologic evidence might be sufficient. *See* Lartigue v. R.J. Reynolds Tobacco Co., 317 F.2d 19, 22-23 (5th Cir. 1963); Pritchard v. Liggett & Myers Tobacco Co., 295 F.2d 292, 294-96 (3d Cir. 1961); *Scientific Proof*, *supra* note 19, at 369-73. Litigation about cigarettes, however, has been stifled by warning labels that preclude warranty claims, and by court holdings that until the labels were put on the packages the manufacturers could not have known about the harm cigarettes could cause and thus could not be held liable. *See* W. Prosser, *supra* note 11, § 99, at 660 & nn.82-83; *Scientific Proof*, *supra* note 19, at 369-73.

57. *See, e.g.*, Estep, *Radiation Injuries and Statistics: The Need for A New Approach to Injury Litigation*, 59 Mich. L. Rev. 259, 273-80 (1960); Forgotson, *Liability For Long-Term Latent Effects of Toxic Agents*, 50 A.B.A.J. 142, 142 (1964); Hall & Silbergeld, *supra* note 5, at 442-43; Henderson, *Medical Causation in Products Liability Disease Litigation*, Trial, June 1981, at 53, 55-57; Mobilia & Rossignol, *The Role of Epidemiology in Determining Causation in Toxic Shock Syndrome*, Jurimetrics J., Fall 1983, at 78, 82-86; Riley, *Toxic Shock Syndrome: Proving Causation Before Science Has*, 6 Am. J. Trial Advoc. 15, 19 (1982); Rosenberg, *supra* note 5, at 856-57, 869-74; Seltzer, *Personal Injury Hazardous Waste Litigation: A Proposal for Tort Reform*, 10 B.C. Envtl. Affairs L. Rev. 797, 815-21,

and a few courts have acknowledged the need to infer causation from comparisons between populations.[58] To date, however, neither commentators nor courts have provided guidance on how to mesh law and epidemiology in a consistent way.

A series of New York cases exemplifies both current developments and current confusion. In *Miller v. National Cabinet Co.*,[59] the New York Court of Appeals reversed an award of workers' compensation

---

846-49 (1982-1983); *Tort Actions for Cancer*, *supra* note 5, at 857. *But see* Dickson, *Medical Causation by Statistics*, 17 Forum 792, 799-808 (1983) (noting shortcomings in the use of epidemiologic evidence); Dore, *A Commentary on the Use of Epidemiological Evidence in Demonstrating Cause-in-Fact*, 7 Harv. Envtl. L. Rev. 429, 431 (1983) ("Because of the confusing and complex nature of epidemiologic evidence, courts should . . . [limit] the use of such evidence as proof of causation . . . .").

The proponents of epidemiology give little guidance on how courts should use it, and except for Forgotson, none address the idea of requiring epidemiologic evidence. A number of commentators seem to have the impression that courts tend not to accept epidemiologic evidence. *See, e.g.*, Rosenberg, *supra* note 5, at 857-58, 869-74; Seltzer, *supra*, at 821-24; *Tort Actions for Cancer*, *supra* note 5, at 848; *see also* Trauberman, *supra* note 3, at 198 (author knows of no case in which an award has been based solely on epidemiologic evidence). Research, however, reveals no case in which a court has held against a plaintiff who has produced evidence sufficient to satisfy the standard proposed in this Article. A large part of the problem is that without a substantive standard, plaintiffs do not know how to present their cases. *Cf.* Schaefer v. Texas Employer's Ins. Ass'n, 612 S.W.2d 199, 205 (Tex. 1980) (plaintiff lost appeal because he failed to produce tests or data).

58. Traces of epidemiologic reasoning have appeared in a variety of cases. *See, e.g.*, Mahoney v. United States, 220 F. Supp. 823, 838 (E.D. Tenn. 1963) (court found for the defendant because there was only a 1 in 24 chance that the plaintiff's leukemia had been caused by radiation), *aff'd*, 339 F.2d 605 (6th Cir. 1964); Braden v. City of Hialeah, 177 So. 2d 235, 236 (Fla. 1965) (per curiam) (plaintiff's claim rejected because she did not show that workplace exposure to sun made probability of contracting skin cancer greater than that of persons with normal exposure to sun); Miller v. Olin Mathieson Chem. Corp., 398 S.W.2d 472, 472-73 (Ky. 1965) (plaintiff's claim rejected because physician's theory of chemical causation of leukemia contradicted by statistical data showing that the incidence of leukemia increased when presence in atmosphere of chemical compounds decreased); Miller v. National Cabinet Co., 8 N.Y.2d 277, 283-84, 168 N.E.2d 811, 814, 204 N.Y.S.2d 129, 133-34 (reference to need for medical statistics showing correlation between exposure to benzol and incidence of leukemia), *modified on other grounds*, 8 N.Y.2d 1025, 170 N.E.2d 215, 206 N.Y.S.2d 796 (1960); Collins v. National Aniline Div., 8 A.D.2d 900, 901, 186 N.Y.S.2d 979, 981 (1959) (reference to comparison of incidence rates of bladder cancer among those exposed to carcinogenic compounds and those not so exposed); Parker v. Employers Mut. Liab. Ins. Co., 440 S.W.2d 43, 47-48 (Tex. 1969) (testimony admitted but held not conclusive that persons exposed to radiation have a higher incidence rate of cancer than non-exposed persons); Ehman v. Department of Labor & Indus., 33 Wash. 2d 584, 595, 206 P.2d 787, 797 (1949) (court held for defendant because plaintiff could not show that but for his employment, he would not have contracted leukemia).

59. 8 N.Y.2d 277, 168 N.E.2d 811, 204 N.Y.S.2d 129, *modified on other grounds*, 8 N.Y.2d 1025, 170 N.E.2d 215, 206 N.Y.S.2d 796 (1960).

benefits to the widow of a worker whose death from leukemia had allegedly been caused by exposure to benzene (also known as benzol). The plaintiff's principal expert witness testified that the incidence of leukemia "is quite high in patients who have been exposed to benzol," and that "it is *possible* that this man's leukemia resulted from his alleged exposure to inhalation of benzol or benzene."[60] In holding for the defendant, the court relied principally on the possibility-probability distinction.[61] It pointed out, however, that "[t]he only possible basis for drawing an inference in favor of claimant . . . would be statistics indicating that in many instances leukemia follows benzol exposure without knowing why."[62]

The allusion in *Miller* to the consideration of statistics as a factor in the determination of causation represents a small step forward in toxic tort theory. Subsequent decisions in New York, however, have not furthered the development of this concept. Most opinions have been couched in terms similar to the plaintiff's argument in *Miller* and have failed to employ statistical data in arriving at their conclusions about causation.[63] In one case, decided for the plaintiff, the expert testified only that he knew at least some of the causes of the disease in question, and that the plaintiff had been exposed to one of them.[64] Two experts in another case stated, with little quantification, that people in the plaintiff's occupation ran a high risk of developing papillary tumors.[65] Again the plaintiff prevailed, though the facts were hardly distinguishable from those in *Miller*. In still another case, the court explicitly found that the statistical requirement had been met, only to be

---

60. *Id.* at 282, 168 N.E.2d at 813, 204 N.Y.S.2d at 132 (emphasis added).

61. *Id.* at 282-83, 168 N.E.2d at 813, 204 N.Y.S.2d at 132-33.

62. *Id.* at 283, 168 N.E.2d at 814, 204 N.Y.S.2d at 133. How to use statistics and how to incorporate other information in drawing biological inferences remained unexplained, though the decision hinted that an eleven-fold increase in the incidence rate in an exposed population might not support a plaintiff's verdict. *Id.* at 285, 168 N.E.2d at 815, 204 N.Y.S.2d at 135.

63. *E.g.*, Shannon v. Grumman Aircraft, 29 N.Y.2d 786, 787-88, 277 N.E.2d 190, 190-91, 327 N.Y.S.2d 71, 72 (1971), *rev'g* 35 A.D.2d 230, 315 N.Y.S.2d 172 (1970); Boney v. Gouverneur Talc Co., 77 A.D.2d 702, 702, 430 N.Y.S.2d 399, 399 (1980); Smith v. Humboldt Dye Works, 34 A.D.2d 1041, 1042, 312 N.Y.S.2d 612, 614 (1970); Benenati v. Tin Plate Lithographing Co., 29 A.D.2d 805, 806, 287 N.Y.S.2d 528, 530 (1968); Amoroso v. Tubular & Cast Prods. Mfg. Co., 17 A.D.2d 1003, 1003-04, 233 N.Y.S.2d 909, 910-11 (1962), *aff'd*, 13 N.Y.2d 992, 194 N.E.2d 694, 244 N.Y.S.2d 787 (1963); Hassell v. Oxford Filing Supply Co., 16 A.D.2d 534, 536, 230 N.Y.S.2d 866, 868 (1962); *see, e.g.*, Yannon v. New York Tel. Co., 86 A.D.2d 241, 244, 450 N.Y.S.2d 893, 895 (1982); Berman v. Werman & Sons, 14 A.D.2d 631, 631, 218 N.Y.S.2d 315, 316 (1961).

64. Benenati v. Tin Plate Lithographing Co., 29 A.D.2d 805, 806, 287 N.Y.S.2d 528, 530 (1968).

65. Smith v. Humboldt Dye Works, Inc., 34 A.D.2d 1041, 1042, 312 N.Y.S.2d 612, 614 (1970).

reversed by the court of appeals, which found "no observable or acceptable correlation between exposure . . . and [disease]."[66] The decisions at both levels fail to indicate the standard by which statistical inference should be judged, or how biological inference should follow from statistics.

The *Miller* line of cases typifies the haphazard way in which courts have addressed the use of comparisons between exposed and unexposed populations to establish toxic tort causation.[67] No clear standard has yet emerged to determine when data and analysis are legally sufficient, or if statistical and non-statistical evidence have been properly integrated.[68] This has clouded legal analysis as well as factfinding.

C. *The More-Likely-Than-Not Test in Toxic Tort Cases*

Courts generally have not held that a toxic tort plaintiff bears a lesser burden of proof on the issue of harmfulness than does the traditional tort law plaintiff.[69] In fact, courts have explicitly adopted the preponderance test in a number of cases in which the harmfulness of a substance was at issue. In *Parker v. Employers Mutual Liability*

---

66. Shannon v. Grumman Aircraft, 29 N.Y.2d 786, 788, 277 N.E.2d 190, 191, 327 N.Y.S.2d 71, 72 (1971), *rev'g* 35 A.D.2d 230, 315 N.Y.S.2d 172 (1970).

67. Courts in other states have also touched upon the evidentiary use of statistical inference in determining toxic tort causation. *See, e.g.*, Miller v. Olin Mathieson Chem. Corp., 398 S.W.2d 472, 473 (Ky. 1965); Schaefer v. Texas Employers' Ins. Ass'n, 612 S.W.2d 199, 201 (Tex. 1981); Parker v. Employers Mut. Liab. Ins. Co., 440 S.W.2d 43, 49 (Tex. 1969); Garner v. Hecla Mining Co., 19 Utah 2d 367, 370, 431 P.2d 794, 796 (1967). *Garner* is the most interesting case because it involved the question how statistical evidence should mesh with non-statistical considerations, one of the principal concerns of epidemiology. A widow appealed the denial of workers' compensation benefits for the death of her husband, who had been a uranium miner. The widow introduced autopsy results showing that her husband's body had contained 34 times as much radioactive lead as the average non-miner's. She also introduced data indicating a high incidence of lung cancer in uranium miners. The court did not find this proof necessarily insufficient, but held that such evidence did not compel an award of benefits. *Id.* at 370, 431 P.2d at 796. The court noted that other factors might have caused the disease, specifically mentioning the fact that the decedent had smoked for approximately twenty years. *Id.* at 371, 431 P.2d at 796-97.

68. Only in a few of the cases that grew out of the 1976 swine flu inoculation program have courts made further progress, but the circumstances surrounding those cases were unique. The increased risk of Guillan-Barre Syndrome (GBS) related to swine flu shots lasted for only a few weeks. Most toxic tort risks are less reversible. Also, because of the number of people involved in the swine flu program and the careful monitoring of it by the Center for Disease Control, very good epidemiologic data were available. See *infra* pt. IV(B)(3).

69. The plaintiff must produce "proof which leads the jury to find that the existence of the contested fact is more probable than its nonexistence." E. Cleary, *supra* note 10, § 339, at 794. See *supra* note 10.

Insurance Co.,[70] for example, the plaintiff alleged that his cancer had been caused by workplace exposure to radiation. He was unsuccessful because he could only establish a low level of exposure, which merely suggested the possibility of causation. The court held that "a possible cause only becomes 'probable' when in the absence of other reasonable causal explanations it becomes *more likely than not* that the injury was a result of its action."[71]

In *McEwen v. Ortho Pharmaceutical Corp.*,[72] the plaintiff claimed that her blindness had been caused by birth control pills. The Oregon Supreme Court upheld her jury verdict, finding that the medical testimony had at least established that the inference of causation was "more probably correct than incorrect."[73] Other toxic tort decisions have been similarly based on the more-likely-than-not test,[74] but except in a few of the swine flu cases,[75] none has come close to considering either the need for epidemiologic evidence or how to analyze such evidence to insure that legal requirements are met.

With the dramatic increase in litigation over latent effects of toxic exposures, the failure to fit known facts into a legal context makes the need for a substantive evidentiary standard ever more pressing. The formulation of a test that will meet this need requires a basic understanding of the philosophy and methods of epidemiology. Properly used and evaluated, epidemiologic evidence will enable courts to adhere to both tort law and scientific principles.

II. Epidemiologic Principles

The elucidation of the relationship between a disease and a factor (e.g., a toxic substance) suspected of causing it lies within the domain of epidemiology.[76] The epidemiologist examines this relationship in the context of populations, comparing the disease experiences of people exposed to the factor with those not so exposed.[77] Although the epidemiologist utilizes statistical methods, the ultimate goal is to draw a biological inference concerning the relationship of the factor to the

---

70. 440 S.W.2d 43 (Tex. 1969).

71. *Id.* at 47 (emphasis added).

72. 270 Or. 375, 528 P.2d 522 (1974).

73. *Id.* at 415 n.36, 528 P.2d at 541 n.36.

74. Sheptur v. Procter & Gamble Distrib. Co., 261 F.2d 221, 224 (6th Cir. 1958) (per curiam); Coburn v. North American Refractories Co., 295 Ky. 566, 174 S.W.2d 757 (1943); Grinnell v. Charles Pfizer & Co., 274 Cal. App. 2d 424, 435, 79 Cal Rptr. 369, 374-75 (1969).

75. See *infra* pt. IV(B)(3).

76. *See* Last, *Scope and Methods of Prevention*, in Maxcy-Rosenau Public Heath and Preventive Medicine 7-8 (J. Last ed. 1980).

77. *See* A. Lilienfeld & D. Lilienfeld, Foundations of Epidemiology 3 (2d ed. 1980).

disease's etiology and/or to its natural history.[78] Stated more formally, "epidemiology can be regarded as a sequence of reasoning concerned with biological inferences derived from observations of disease occurrence and related phenomena in human population groups."[79] It is an integrative, eclectic science utilizing concepts and methods from other disciplines, such as statistics, sociology and demography for the study of disease in populations.

To understand epidemiologic methods and reasoning, one must understand how epidemiology grew out of its component disciplines. The natural philosophers of the seventeenth century initiated a method of reasoning based on the premise that one can mathematically model a population's mortality experience.[80] This work developed into the modern fields of demography, vital statistics, and subsequently, epidemiology.

One of the tools that these scientists developed was the life table, known until the early 1900's as the "table of mortality."[81] The first life tables reflected only the aggregate mortality experience in a population.[82] They provided no record of individual diseases because the concept of specific diseases had not yet crystallized.[83] Indeed, al-

---

78. Lilienfeld, *Definitions of Epidemiology*, 107 Am. J. Epidemiology 87, 89 (1978).

79. A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 4.

80. *See* Lilienfeld, *"The Greening of Epidemiology": Sanitary Physicians and the London Epidemiological Society (1830-1870)*, 52 Bull. Hist. of Med. 503, 504 (1979); Lorimer, *The Development of Demography*, in The Study of Population 124, 127 (P. Hauser & O. Duncan eds. 1959).

81. *See* Lilienfeld & Lilienfeld, *The French Influence on the Development of Epidemiology*, in Times, Places and Persons: Aspects of the History of Epidemiology 28, 28 (A. Lilienfeld ed. 1980). Figure I shows a typical life table, a tabulation of a given population's mortality experience.

FIGURE 1

A TYPICAL LIFE TABLE

Age	Population at Start of Age	Deaths
0-1	1,000	20
1-4	980	80
5-14	900	250
15-24	650	250
25-34	400	300
35 and over	100	100

82. J. Farren, Historical Essay on the Rise and Early Progress of the Doctrine of Life-Contingencies in England (London 1844); J. Francis, Annals and Legends of Life Assurance 87-97 (London 1853); *see* Lilienfeld & Lilienfeld, *supra* note 81, at 28.

83. *See* Temkin, *Comment on Hilt's "Epidemiology and the Statistical Movement,"* in Times, Places and Persons: Aspects of the History of Epidemiology 61 (A. Lilienfeld ed. 1980).

though the notion of statistically viewing the mortality experience of a population dates from the mid-1600's, not until the 1800's did the concept of disease specificity emerge. This development permitted scientists to make accurate correlations and to draw meaningful causal inferences.[84]

A. *The Definition of Disease*

Although concern about the exact definition of a disease began with communicable diseases, it is of equal concern when dealing with chronic diseases such as cancer, heart disease and stroke. The epidemiologist must begin his investigation with a clear, precise definition of the disease being studied.[85] Within the medical community, disease is viewed as an entity characterized by at least two of the following criteria: "a recognized etiologic agent (or agents); an identifiable group of signs and symptoms; [and/or] consistent anatomical alterations [that is, lesions or a pathologic state being present]."[86] This definition of disease does not differ markedly from that used by lawyers: "An illness or an abnormal state having a definite pattern of symptoms."[87] Neither statement, however, suffices for an epidemiologic investigation, which requires an exact definition of the disease being studied.

The definition of a particular disease depends on its nature and must be sufficiently precise to permit exclusion of all other diseases from consideration. The "gold standard" definition is that of the pathologist, as it is based on the histologic characteristics of the disease. For diseases defined by pathophysiologic changes, such as asthma, other characteristics, such as physiologic ones, may be used. Some diseases and syndromes, such as volvulus,[88] are best defined in terms of what is observed during surgical intervention. The internist

---

84. The melding of the concepts of statistics and specificity was accomplished in Paris and London in the mid-nineteenth century by Pierre Charles Alexandre Louis and his English students. Louis' investigations of typhus, typhoid fever and tuberculosis are still considered classics in both epidemiology and clinical medicine. His insistence on accurate data remains a keystone of sound epidemiologic work. *See* A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 31 n.7. Louis and his students were concerned with the specificity of disease, i.e., a precise definition of the disease which excludes all other diseases from consideration. *See* Temkin, *supra* note 83, at 61.

85. A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 134-35.

86. Stedman's Medical Dictionary 401 (23d ed. 1976).

87. Black's Law Dictionary 420 (5th ed. 1979).

88. Volvulus is one form of intestinal blockage in which the intestine twists upon itself, thereby causing an obstruction. As the lesion in this condition is grossly visible upon surgical entry into the abdominal cavity, the surgeon can readily ascertain the pathology upon such intervention. Indeed, attempting to define this condition based on its histology is nearly impossible due to the macroscopic nature of its pathology.

seeks to relate these histologic and/or other characteristics to the clinical signs and symptoms exhibited by affected patients. The patient's disease is thereby diagnosed.

Because the epidemiologist depends on laboratory tests and those clinical signs and symptoms noted by the clinician, he needs a measure of the accuracy of these clinical indicators as they relate to the definition of the disease. The two most commonly used measures of the accuracy of clinical diagnoses are "sensitivity" and "specificity."[89] "Sensitivity" is defined as the proportion of correct diagnoses as ascertained by clinical signs or symptoms and/or laboratory tests of those afflicted with the disease. The percentage of instances in which the disease is not so diagnosed when it is in fact absent is known as "specificity."[90]

To determine the sensitivity and specificity of a particular clinical diagnosis or laboratory test, the epidemiologist selects individuals known to have or not to have the disease, then applies the test to these individuals. If either sensitivity or specificity is low, the quality of the epidemiologist's data is correspondingly diminished.

B. *Determining the Relationship between Incidence of Disease and Exposure to a Factor*

Once the epidemiologist has defined the disease of interest, he seeks to compare the rate of disease development (incidence rate) among

89. A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 150.
90. The following figure illustrates these concepts:

FIGURE 2

INDICES TO EVALUATE THE ACCURACY OF A TEST OR DIAGNOSTIC EXAMINATION: SENSITIVITY AND SPECIFICITY

Test or Examination	Disease Present	Disease Absent
Positive (Indicating disease is probably present)	A (true positives)	B (false positives)
Negative (Indicating disease is probably absent)	C (false negatives)	D (true negatives)
Totals	A + C	B + D

Sensitivity is defined as the percent of those who have the disease, and are so indicated by the test. Thus,

$$\text{Sensitivity (in percent)} = \frac{A}{A + C} \times 100$$

Specificity is defined as the percent of those who do not have the disease and are so indicated by the test. Thus,

$$\text{Specificity (in percent)} = \frac{D}{B + D} \times 100$$

those exposed to the factor of interest with the rate among those not so exposed. The incidence rate is a measure of the probability that an individual will develop the disease. Hence, the epidemiologist is interested in determining if exposure to the factor changes the probability that an individual will develop the disease.[91] If there is a gradation in the degree of exposure, the possibility of a corresponding gradation in incidence rates exists and merits investigation. The two principal approaches to collecting and analyzing morbidity/mortality data for exposed and non-exposed individuals are the demographic study and the epidemiologic study. In the former, the subjects within the two groups are viewed in the aggregate, while in the latter the subjects are viewed individually.[92] The results of demographic studies are used to generate etiologic hypotheses, which are then tested through epidemiologic studies.

1. The Demographic Study

Demographic studies explore either morbidity, if the investigator seeks to explain sickness, or mortality, if the investigator seeks to explain death. In either case, a study initially seeks to determine the accuracy and completeness of the statistics being analyzed and then attempts to ascertain how such statistics are related to possible etiologic factors, such as age, sex, cigarette consumption or asbestos exposure. One might, for example, examine the relationship between annual asbestos use in the United States from 1910-1950 and the annual mortality rates for mesothelioma in the United States from 1940-1980. Before drawing conclusions from the relationship between asbestos exposure and mesothelioma, however, the epidemiologist must determine the accuracy of the available mortality and exposure data in order to ensure that there has not been under- or over-reporting of either asbestos use or mesothelioma mortality. Studies have indicated that such data are available and accurate and that there is a positive relationship between asbestos use and mortality from mesothelioma.[93] Although such a positive correlation is supportive of a possible causal relationship between the two, it is by no means conclusive.[94]

No matter how compelling the findings in a demographic study, it must be recognized that such observations refer to groups and not to the individuals within the groups. A correlation may exist between a

---

91. *See* A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 14, 191.

92. *See id.* at vii, 191-94.

93. National Cancer Institute, National Institute of Environmental Health Sciences & National Institute for Occupational Safety and Health, Estimates of the Fraction of Cancer in the United States Related to Occupational Factors 8-11 (1978) [hereinafter cited as Occupational Factors].

94. *See* Goodman, *Ecological Regressions and Behavior of Individuals*, 18 Am. Soc. Rev. 663, 663 (1953); Robinson, *Ecological Correlations and the Behavior of Individuals*, 15 Am. Soc. Rev. 351, 351-52, 357 (1950).

factor and the incidence of a disease even though no causal relationship exists. The classic example of this phenomenon is the linear relationship between pig iron production in the United States and the birth rate in Great Britain.[95] Clearly, such an association is spurious. This problem is known as an "ecological fallacy," and it imposes an inherent limitation on the use of demographic studies in inferring a causal relationship between a factor and a disease.[96] Demographic studies are used mainly to focus attention on a possible association between a factor and a disease, the elucidation of which requires further, more refined modes of study. In order to demonstrate the association in terms of the individual members of a group, the investigator utilizes the epidemiologic study.[97]

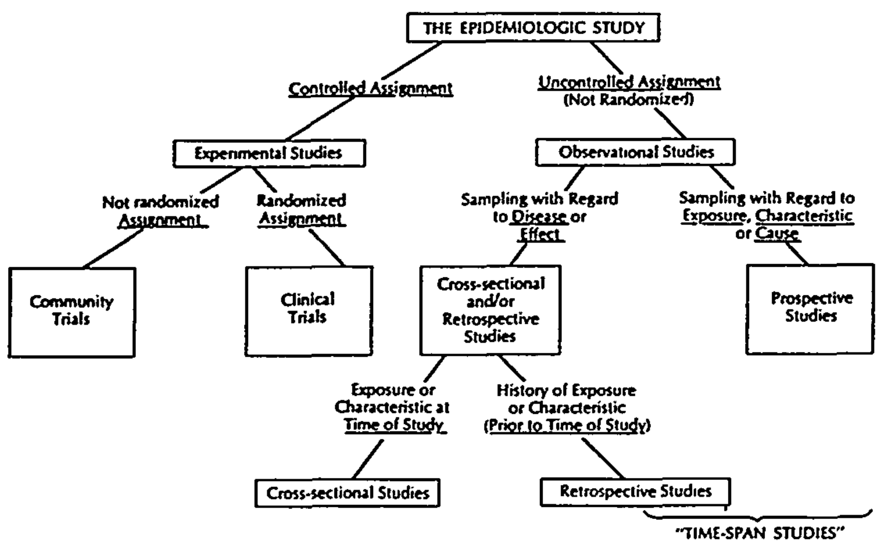
2. The Epidemiologic Study

The epidemiologic study attempts to explore and clarify a possible association between a factor and a disease within individuals in a population. For epidemiologists, it represents the application of the scientific method to human populations. In the scientific method, the investigator observes the effect of a single modification in the environment of one of two otherwise identical animals. Similarly, in an epidemiologic study, one seeks to observe the effect of exposure to a single factor upon the incidence of disease in two otherwise identical populations.

There are two major types of epidemiologic studies: experimental and observational.[98] In experimental studies, the epidemiologist assigns the exposure status to individuals. If the assignment is not performed randomly, it is termed a "community trial." The use of fluori-

95. G. Snedocor & W. Cochran, Statistical Methods 189 (6th ed. 1967).
96. *See* A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 14.
97. *See id.* at 191.
98. *See id.* at 191-94. Figure 3 depicts the difference between the experimental and the observational study.

FIGURE 3
THE ANATOMY OF THE EPIDEMIOLOGIC STUDY



dation in water to prevent dental caries was tested in this way.[99] If the epidemiologist randomly assigns individuals to exposed and non-exposed groups, the study is a "clinical trial." The purpose of the randomization is to ensure that the only difference between the two groups is in the exposure; and that in all other respects, the groups are comparable.[100] Almost every new drug authorized for use by the Food and Drug Administration has been tested by such a clinical trial. While clinical trials are definitive studies,[101] they are not commonly encountered in toxic tort litigation because it is seldom possible to experiment by assigning individuals to an exposure.

The assignment of exposure, and thus an experimental study, is feasible only when it is ethical. It would be unethical, for instance, to assign individuals to exposure to cigarette smoking. The observational (non-experimental) study is uniquely suited to investigating situations in which controlled assignment is either unethical or difficult to achieve. In observational studies, the epidemiologist systematically observes the disease experience of individuals whose exposure status has been determined by themselves or by others in a nonrandomized manner. One might, for example, be interested in determining the difference in lung cancer incidence between smokers and non-smokers. If the epidemiologist views the population in terms of the individuals' exposure, the study type is "prospective." The investigator first determines if the individuals are cigarette smokers, then follows them over a sufficient number of years to see if their lung cancer incidence rate differs from that of non-smokers. If the epidemiologist views the populations in terms of individual disease status, the study is either "retrospective" or "cross-sectional." Retrospective studies focus on past exposure while cross-sectional studies consider current exposure. The investigator selects individuals who have or do not have lung cancer and then determines whether or not they are or have been cigarette smokers.

a. *Prospective Studies*

The prospective study is a powerful way to investigate the relationship between a factor and a disease because it closely approximates the classical scientific method. The investigator identifies two populations (or representative samples thereof), one composed of individuals who have been exposed to the factor and one of individuals who have not been so exposed.[102] Ideally, these populations will be otherwise identi-

---

99. *See id.* at 5-6.
100. *See id.* at 257.
101. *See id.* at 256-57; D. Schwartz, R. Flemant & J. Lellouch, Clinical Trials (M. Healy trans. 1980).
102. A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 226; *see* J. Schlesselman, Case-Control Studies 14-15 (1982).

cal.[103] The investigator follows these populations for a period of time (possibly many years), observing the incidence rates of disease in each population. If the two groups are comparable, any difference in disease incidence can then be related either to the factor or to the sampling process, that is, to chance. Several statistical methods are available for assessing whether a difference in incidence rates results from sampling rather than from exposure to the factor.[104] After eliminating chance and determining that a statistically significant relationship between the disease and the factor exists, the epidemiologist's next task is to estimate the magnitude of the association. The accepted means of measuring such an association is the calculation of the

---

103. If the two groups are not in fact comparable, statistical methods have been developed for adjusting the relative risk to account for the differences between them. J. Fleiss, Statistical Methods for Rates and Proportions 237-55 (2d ed. 1981); *see* Cochran, *Some Methods for Strengthening the Common* X^2 Tests, 10 Biometrics 417 (1954); Mantel & Haenszel, *Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease*, 22 J. Nat'l Cancer Inst. 719, 730 (1959).

104. *See* P. Armitage, Statistical Methods in Medical Research (1971); J. Fleiss, *supra* note 103. In both books, every chapter relates in some way to how statistical studies should be performed, but of particular interest on the question of sampling are chapters 3 and 4 in Fleiss and chapter 6 in Armitage.

The importance of statistical significance testing is that it enables the investigator to determine if the difference observed between two samples represents a true difference between the populations or if it is instead the result of the sampling process. *See* D. Barnes, Statistics as Proof—Fundamentals of Quantitative Evidence 143-45 (1983). *See generally* I. Hacking, Logic of Statistical Inference (1965).

The investigator will usually state the hypothesis that there is no actual difference as the "null hypothesis." For example, in a study examining the mortality of cigarette smokers compared to that of non-smokers, the null hypothesis (H_o) would be that the mortality rates for both groups are the same, and thus that cigarette smoking has no impact on mortality (the status quo). Alternatively, the null hypothesis can be viewed as the statement that the investigator is seeking to disprove. In either case, the conjugate of H_o is H_1 (also termed H_a). The statistical significance test provides the probability that the observed difference is due to chance if H_o is, in fact, true. If that probability is sufficiently small (5% being the most-commonly used level), then the investigator "rejects" H_o, concludes that its conjugate, H_1, is true, and completes his investigation using H_1 as an established fact. (This analysis is sometimes done using confidence intervals, which are fully equivalent to significance tests.)

It should be noted that with a sufficient number of observations from each population, a statistically significant result will be observed for even very small differences, which may represent little or no biological difference. It should also be noted that the statistical significance test does not have anything to do with the evaluation of the remainder of the investigation. The determination of the probability of the observed events being attributable to random events, that is, secondary to the sampling process, does not in fact assign a probability level to the results of the investigation being "correct." Once the investigator has determined that the differences he has observed are not in fact the result of random chance, he has made his inference as far as the statistical significance tests are concerned, and he then goes on to complete the remainder of his investigation, including the determination of biological inferences, without recourse to the probability figure that he derived in conducting the statistical significance test.

relative risk, which is the ratio of the incidence rate of disease in the exposed group divided by that rate in the non-exposed "control" group.[105] If there is no association between the factor and the disease, the relative risk is 1.0; that is, the incidence rates for the exposed and non-exposed groups are equal.

The greater the magnitude of the observed relative risk, the stronger the association between the factor and the disease. If the factor were the only cause of the disease, the relative risk would be infinite because the incidence of disease in the unexposed group would be zero. Because most diseases have multi-factorial etiologies, however, it is rare to observe a relative risk greater than 10. When a relative risk of 10 or more is observed, one can be reasonably certain that it represents a causal relationship. For example, the relative risk for mesothelioma from asbestos exposure, which is widely recognized as causal, is between 50 and 80.[106] By comparison, the relative risk for leukemia in children who have been irradiated *in utero* is only 1.6 times that of children who were not so irradiated.[107] This represents a relatively small increase in the risk of developing leukemia for the irradiated children, which reflects a relatively weak causal relationship.

---

105. *See* J. Fleiss, *supra* note 103, at 64-65; A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 209; Cornfield, *A Method of Estimating Comparative Rates from Clinical Data: Applications to Cancer of the Lung, Breast and Cervix*, 11 J. Nat'l Cancer Inst. 1269, 1269 (1951); Mantel & Haenszel, *supra* note 103, at 730. See Figure 4.

FIGURE 4

EXAMPLE OF COMPUTATION OF RELATIVE RISK

1. Groups A and B are assumed identical except for exposure to Factor F. (If not identical, there are methods of adjustment that still allow valid comparisons).
2. Incidence of disease D in Group A (exposed to Factor F) is 50 per 100,000 population. Incidence of the disease in Group B (not exposed) is 5 per 100,000.
3. Relative risk (r) of exposed to non-exposed is 50/5 = 10.0.

106. Love, *Biological Aspects of Associations Between Environmental Exposures and Cancer*, 37 Am. Statistician 413, 417 (1983).

107. Lilienfeld, *Epidemiology of Infectious and Non-Infectious Disease: Some Comparisons*, 97 Am. J. Epidemiology 135, 141 table 3 (1973). It should be noted that this relative risk was estimated from data collected in a retrospective study. It is presented as an illustration of the importance of the magnitude of the relative risk in making epidemiologic/biological inferences.

The prospective study, although very reliable, is difficult and expensive to conduct. It is not always possible to identify populations that are exposed and not exposed to a factor. Frequently, the epidemiologist is unable to follow the two groups for the period of time required. Hence, epidemiologists have developed and extensively used the retrospective study.

b. *Retrospective Studies*

Whereas a prospective study investigates the disease experience of exposed and non-exposed groups, the epidemiologist performing a retrospective study begins with individuals who already have (cases) or do not have (controls) the disease under investigation.[108] He then determines whether or not each individual has a past exposure to the factor, presumably prior to the onset of the pathologic process resulting in the disease. Cases are usually ascertained in a hospital setting. Control groups are commonly selected in several different ways, including: (1) "hospital controls," in which hospital patients who are not cases, but have different diseases, serve as controls;[109] (2) "population" or "neighborhood controls," in which a random sample of the case's neighbors or other similar groups constitutes the controls;[110] and (3) "matched population" or "matched neighborhood controls," in which population or neighborhood controls are matched to the cases so that various factors known or suspected to be unrelated to the disease are similarly distributed in the case and the control populations.[111] The retrospective study is inherently limited because one cannot directly ascertain disease incidence rates among the exposed and non-exposed groups; hence, the relative risk cannot be calculated directly.[112] There is, however, a statistic known as the "odds ratio"[113] that approximates the relative risk in those instances in which the disease incidence rate in the non-exposed population is low. As the odds ratio increases, so does the relative risk.

Retrospective studies in which hospital controls are used, unlike prospective studies, may be subject to a major bias in the selection of the controls, known as a "Berksonian bias."[114] The bias results from

---

108. A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 194; *see The Case Control Study: Consensus and Controversy*, 32 J. Chronic Diseases 1 (1979).

109. *See* A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 196-97 & table 8-4.

110. *See id.* at 197 table 8-4.

111. *See id.* at 197-98 & table 8-4.

112. Cornfield, *supra* note 105, at 1269.

113. *See* Fleiss, *Confidence Intervals for the Odds Ratio in Case-Control Studies: The State of the Art*, 32 J. Chronic Diseases 69 (1979).

114. A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 202; *see* Berkson, *Limitations of the Application of Fourfold Table Analysis to Hospital Data*, 2 Biometrics 47, 49-51 (1946).

the differing probabilities of admission into the hospital for cases and hospital controls. If the probabilities of admission for each of these two groups are equivalent, there is no Berksonian bias.[115] The maximum increase in the observed odds ratio that a Berksonian bias usually produces in the absence of any relationship between a factor and a disease is approximately three.[116] Hence, if an odds ratio is observed to be greater than three, it is unlikely to have resulted entirely from the operation of a Berksonian bias.

c. *Cross-Sectional Studies*

Epidemiologic studies usually are concerned with relating antecedent exposure with subsequent disease occurrence. There are, however, occasions when the epidemiologist is interested in determining the relationship between current exposure and current disease status. This association can be elucidated by the cross-sectional study.[117] However, as diseases involved in toxic tort litigation generally have significant latency periods, cross-sectional studies are usually of little use in determining causation.

d. *Attributable Risk*

Observational studies are all directed at determining the relative risk of developing a disease that is associated with exposure to a factor. The relative risk, however, expresses only the magnitude of that association.[118] The statistical measure of a factor's relationship to a disease in the population is the "attributable risk."[119] It was originally described as the percentage decline in the population's disease incidence that would occur if the population's exposure to the factor were eliminated.[120] For example, the risk of lung cancer attributable to smoking in the United States today is approximately eighty percent. In other words, if smoking were eliminated in the United States, the incidence of lung cancer would decline by about eighty percent.[121]

---

115. *See* A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 199-202.

116. Lilienfeld, *The Maximum Relative Risk Produced by a Berksonian Bias* (unpublished manuscript 1983) (available in files of *Fordham Law Review*). *See* A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 201-02.

117. A cross-sectional study is identical to a retrospective one except that the investigator is concerned with current exposure status. Therefore, it shares the retrospective study's limitation in estimating relative risks.

118. A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 217-18, 302.

119. *Id.* at 217.

120. Walter, *Calculation of Attributable Risks from Epidemiological Data*, 7 Int'l J. Epidemiology 175, 175 (1978); *see* Levin, *The Occurrence of Lung Cancer in Man*, 9 Acta Unio Internationala Contra Cancrum 531, 536 (1953).

121. A. Lilienfeld, Foundations of Epidemiology 256 (1st ed. 1976).

Alternatively, the attributable risk may be viewed as representing the proportion of the disease that is statistically attributable to the factor.[122] Using the example of lung cancer and cigarette smoking, one could say that cigarette smoking accounts for approximately eighty percent of the incidence of lung cancer in the United States. The attributable risk, therefore, is a composite measure that takes into account both the relative risk of disease if exposed and the proportion in the population so exposed.[123] It is an essential tool in examining the sufficiency of epidemiologic evidence.

---

122. Walter, *The Distribution of Levin's Measure of Attributable Risk*, 62 Biometrika 371, 371 (1975).

123. From the equation in Figure 5, it can be seen that for the attributable risk to be high for a given factor (i.e., greater than 0.5), both the relative risk (r) and the proportion in the population so exposed (b) must be relatively large.

FIGURE 5

CALCULATION OF ATTRIBUTABLE RISK

$$\text{Attributable Risk} = \frac{b\,(r-1)}{b\,(r-1)+1}$$

b = proportion of total population exposed to factor
r = relative risk

The table in Figure 6 shows how attributable risk varies within these parameters. If an investigator restricts the definition of exposure, thereby increasing the relative risk, the proportion of exposed people in the population would be lower and the attributable risk would remain approximately the same.

FIGURE 6

ATTRIBUTABLE RISKS AS A PROPORTION FOR SELECTED VALUES OF RELATIVE RISK AND PROPORTION OF POPULATION WITH THE CHARACTERISTIC*

b = Proportion of Population with Characteristic (percent)	r = Relative Risk			
	2	4	10	12
10	.09	.23	.47	.52
30	.23	.47	.73	.77
50	.33	.60	.82	.84
70	.41	.67	.86	.89
90	.47	.73	.89	.91
95	.49	.74	.90	.92

$$^{*}\text{Attributable Risk} = \frac{b\,(r-1)}{b\,(r-1)+1}$$

C. *Biological Inferences from Epidemiologic Data*

Demographic and epidemiologic studies both facilitate the elucidation of the statistical association between a factor and a disease. In order to draw the biological inference that a causal relationship exists, however, the epidemiologist must integrate additional scientific information. The derivation of such an inference requires rigorous consideration of laboratory, experimental, demographic and epidemiologic data.[124]

A causal inference must be biologically plausible and must conform to generally accepted theories. With the advent of the germ theory, criteria for determining whether a given bacteria caused a disease became necessary. Thus the Henle-Koch Postulates, developed in the nineteenth century, permitted the inference that a given species of bacteria, such as *Vibrio cholera*, is the etiologic agent of a given disease, such as Asiatic cholera. These postulates were:

1. The organism must be found in all cases of the disease in question.
2. It must be isolated from patients with the disease and grown in pure culture.
3. When the pure culture is inoculated into susceptible animals or man, it must reproduce the disease.[125]

The success of epidemiology in elucidating the relationship between non-bacterial causes of disease in the 1930's to the 1950's necessitated extension of the Henle-Koch Postulates in order to to derive biological inferences about the relationship between a factor and a disease.[126] Much of the initial work on these modifications was conducted with a view to establishing the relationship between cigarette smoking and lung cancer. As the breadth of epidemiology expanded, these ideas were generalized. They have been stated formally by Evans[127] and are

124. It should be noted that it is possible to have an inadequately developed biological inference regarding the relationship between a factor and a disease, yet still have a statistically plausible relationship. *See* A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 315-16. The necessary biological knowledge may not be available at the time that the statistical association is found. An example of this occurrence is the relationship between oral contraceptives and various circulatory diseases. *Id.* at 315-16. When an association was discovered, there was no laboratory evidence to support a causal inference. However, the statistical association provided direction for laboratory workers in their research. The resulting laboratory data provided the necessary biological facts for the causal relationship to be stated. *Id.* at 316.

125. *Id.* at 292.

126. For the purposes of this Article, the following definition of a causal relationship will be used: "A causal relationship would be recognized to exist whenever evidence indicates that the factors form part of the complex of circumstances that increases the probability of the occurrence of disease and that a diminution of one or more of these factors decreases the frequency of that disease." *Id.* at 295.

127. Evans, *Causation and Disease: The Henle-Koch Postulates Revisited*, 49 Yale J. Biology & Med. 175 (1976).

now known as the Henle-Koch-Evans Postulates. Widely accepted by epidemiologists as the valid criteria for arriving at biological etiological inferences,[128] the postulates are:

1. The prevalence rate of the disease should be significantly higher in those exposed to the hypothesized cause than in controls not so exposed (the cause may be present in the external environment or as a defect in host responses).
2. Exposure to the hypothesized cause should be more frequent among those with the disease than in controls without the disease when all other risk factors are held constant.
3. Incidence of the disease should be significantly higher in those exposed to the cause than in those not so exposed, as shown by prospective studies.
4. Temporally, the disease should follow exposure to the hypothesized causative agent with the distribution of incubation periods on a log-normal-shaped curve.
5. A spectrum of host responses should follow exposure to the hypothesized agent along a logical biologic gradient from mild to severe.
6. A measurable host response following exposure to the hypothesized cause should have a high probability of appearing in those lacking this response before exposure (e.g., antibody, cancer cells) or should increase in magnitude if present before exposure; this response pattern should occur infrequently in persons not so exposed.
7. Experimental reproduction of the disease should occur more frequently in animals or man appropriately exposed to the hypothesized cause than in those not so exposed; this exposure may be deliberate in volunteers, experimentally induced in the laboratory, or demonstrated in a controlled regulation of natural exposure.
8. Elimination or modification of the hypothesized cause or of the vector carrying it should decrease the incidence of the disease (e.g., control of polluted water, removal of tar from cigarettes).
9. Prevention or modification of the host's response on exposure to the hypothesized cause should decrease or eliminate the disease (e.g., immunization, drugs to lower cholesterol, specific lymphocyte transfer factor in cancer).
10. All of the relationships and findings should make biological and epidemiologic sense.[129]

---

128. A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 317-18.

129. *Id.* The first three postulates embody the same concept, that is, that the incidence of disease should be greater in those exposed than in those not exposed for cross-sectional, retrospective and prospective studies. Postulate 4 refers to the epidemic curve, an epidemiologic concept originally developed for infectious diseases that is also applicable to such chronic diseases as cancer. *See* A. Lilienfeld & D. Lilienfeld, *supra* note 77, at 54-56. Postulates 5 and 6 relate to "host responses,"

Satisfaction of these criteria enables the epidemiologist to move beyond a correlation to form a biological inference that is applicable to all contemporary situations. The importance of the last criterion of the Henle-Koch-Evans Postulates cannot be over-emphasized because only its satisfaction can translate statements of statistical associations into inferences understandable within a biological context (concerning a pathophysiological process with a defined cause).

The approach to epidemiologic problems described above is a generally accepted one. Although specific aspects of that approach, such as the extensions made by Evans to the Henle-Koch Postulates, have changed over time, the basic framework of reasoning has remained essentially unaltered since its inception in the nineteenth century. The major change over the past 150 years has not been in the epidemiologic approach to disease problems per se, but rather in the precision and refinement of the methods used to make biological inferences.[130]

III. An Evidentiary Standard Combining the More-Likely-Than-Not Test and Epidemiology

A. *Requirement that Plaintiff Prove that Allegations of Causation Are More-Likely-Than-Not True*

Basic to this Article is the premise that a toxic tort plaintiff bears the burden of proving causation by a preponderance of the evidence.[131] The plaintiff is regarded as the legal aggressor, the one who wants the court to change the present state of affairs.[132] Thus "policy consider-

---

which include such phenomena as fevers, increases in the levels of antibodies to a bacteria or virus, or increases in the number of white cells.

130. Lilienfeld & Lilienfeld, *A Century of Case-Control Studies: Progress?*, 32 J. Chronic Diseases 5, 13 (1979).

131. See *supra* note 10 and accompanying text. The burden of proof encompasses the burden of producing evidence and the burden of persuasion. The former imposes on a party the obligaton to present evidence theoretically sufficient to sustain his version of the facts at issue; the latter determines which side loses if the factfinder is not sufficiently convinced at the end of the trial. E. Cleary, *supra* note 10, § 336, at 783-84; *see* Belton, *Burdens of Pleading and Proof in Discrimination Cases: Toward a Theory of Procedural Justice*, 34 Vand. L. Rev. 1205, 1213 (1981). See *supra* note 10. This Article focuses on the production rather than the persuasion aspect of the burden of proof; its concern is the sufficiency of evidence. The two burdens are conceptually linked, however, because a decision as to whether a party has satisfied the production burden cannot be made without considering the degree of certainty required to meet the persuasion burden. See *infra* pt. IV(B) for a discussion of why the more-likely-than-not test results in the appropriate degree of certainty.

132. Louisell, *Construing Rule 301: Instructing the Jury on Presumptions in Civil Actions and Proceedings*, 63 Va. L. Rev. 281, 285 (1977); *see* Belton, *supra* note 131, at 1213; Cleary, *Presuming and Pleading: An Essay on Juristic Immaturity*, 12 Stan. L. Rev. 5, 7 (1959).

ations of fairness suggest that [he] should be required to prove his claim to relief."[133] While there have been exceptions to this general rule in cases not involving toxic torts,[134] the principles on which the exceptions have been based do not indicate that toxic tort defendants should bear the "burden of disproof" when toxic tort plaintiffs cannot produce sufficient evidence of causation.

Commentators usually discuss reversal of the burden of proof in the context of presumptions, which are created for reasons of policy, fairness and convenience.[135] Judicial analysis, however, usually reduces to evaluation of probabilities and consideration of which party has superior access to proof.[136] Neither of these factors weigh against the typical toxic tort defendant. Consider the presumption that a driver acts in the course of his employment when he drives a vehicle that is owned by his employer. "Although it is known that employees

---

The similarity between civil procedure and the scientific method in dealing with those who seek to change the status quo is also instructive. Both law and science do, after all, strive to determine as nearly as possible what "really" occurs or has occurred, and both have developed means for making decisions in the face of uncertainty. It is interesting that science, like the law, generally insists that a new finding be well-established by evidence before it is accepted as part of the body of scientific knowledge.

133. Belton, *supra* note 131, at 1213.

134. *See, e.g.*, Wells v. Metropolitan Life Ins. Co., 107 Ga. App. 826, 831-32, 131 S.E.2d 634, 638 (1963) (presumption in contract case in plaintiff's favor that her pregnancy extended nine full months); Johnson v. Secretary of State, 406 Mich. 420, 440-42, 280 N.W.2d 9, 14-15 (1979) (presumption of negligence stemming from automobile driver's flight from accident in violation of statute).

135. *See* Belton, *supra* note 131, at 1217; Cleary, *supra* note 132, at 11; James, *Burdens of Proof*, 47 Va. L. Rev. 51, 65 (1961); Louisell, *supra* note 132, at 292-93. *But cf.* Laughlin, *In Support of the Thayer Theory of Presumptions*, 52 Mich. L. Rev. 195, 219 (1953) (balance of probabilities should be used to determine whether plaintiff's burden has been fulfilled). One relatively recent case listed eleven factors to be considered in allocating the burden of proof. Nelson v. Hughes, 290 Or. 653, 658-59, 625 P.2d 643, 645-46 (1981).

Commentators have disagreed about whether presumptions operate to shift both the burden of persuasion and the burden of production, or only the latter. *See* Allen, *Presumptions in Civil Actions Reconsidered*, 66 Iowa L. Rev. 843, 862-67 (1981) (mechanical use of presumptions should be discarded); Hecht & Pinzler, *Rebutting Presumptions: Order Out of Chaos*, 58 B.U.L. Rev. 527, 547-58 (1978) (distinguishing three situations in which presumptions arise); Ladd, *Presumptions in Civil Actions*, 1977 Ariz. St. L.J. 275, 283-88 (questioning whether all presumptions should be treated alike). *Compare* Laughlin, *supra*, at 209-12 (only production burden should be shifted), *with* Morgan, *Presumptions*, 12 Wash. L. Rev. 255, 281 (1937) (both burdens should be shifted).

136. James, *supra* note 135, at 66; *see* International Bhd. of Teamsters v. United States, 431 U.S. 324, 359 n.45 (1977) ("Presumptions shifting the burden of proof are often created to reflect judicial evaluations of probabilities and to conform with a party's superior access to the proof."). *But see* Dworkin, *supra* note 10, at 1161 (policy and fairness are determinative); Laughlin, *supra* note 135, at 219 (only presumptions based on probability are necessary).

sometimes use their employers' vehicles for purely private missions . . . that would constitute a distinct minority of cases."[137] Similarly, because services rendered in the context of a business relationship are not often performed gratuitously, a defendant denying an obligation to pay for such services would have the burden of proving that the obligation did not exist.[138] Such a common sense analysis of what is probable does not support making an exception to the general rule on proof of causation in toxic tort cases. Most diseases, including cancer, do not usually result from tortious conduct, or from exposure to identifiable man-made substances.[139]

Courts have also justified shifting the burden of proof because a defendant has superior access to evidence, but only under unusual circumstances such as when goods are damaged in a bailee's possession.[140] Such circumstances do not exist in most toxic tort cases because the problem encountered in determining causation is not the inaccessibility of evidence, but rather its non-existence or insufficiency. Epidemiologic analysis, the proper basis for recovery, can be performed by either plaintiffs or defendants.[141] A defendant may already possess the necessary records or data for an epidemiologic study, but given sufficient grounds for initiating a suit and a sufficient showing of relevance, discovery rules would make these available to the plaintiff. Thus, neither access to evidence nor probability warrants shifting the burden of proof to defendants in toxic tort cases. Toxic tort plaintiffs should be held to the same requirements as plaintiffs in most other tort actions. They should be required to produce evidence sufficient to establish that the substance at issue more likely than not caused the injury or disease in question.[142]

---

137. Laughlin, *supra* note 135, at 215.

138. E. Cleary, *supra* note 10, § 337, at 787.

139. See *infra* note 220. One court, in holding for a toxic tort defendant, has explicitly noted the lack of such a general relationship between exposure and disease. Miller v. Olin Mathieson Chem. Corp., 398 S.W.2d 472, 473 (Ky. 1966) (noting that while organic chemical usage had increased, the overall incidence of leukemia, the disease at issue, had decreased).

140. James, *supra* note 135, at 66.

141. The only barrier to equal accessibility might be disparity in financial capabilities. No theory, however, would impose on a rich defendant the duty to develop a case for a poor plaintiff.

142. Requiring that a plaintiff sustain the burden of proof by a preponderance of the evidence derives from the practical objective of maximizing the number of cases decided correctly. Unlike criminal law, which is skewed toward avoiding incorrect guilty verdicts, tort law seeks to allocate neutrally the cost of damages or injuries. In most cases its goal is to minimize misallocation, which is best accomplished by using the more-likely-than-not test. *See* Cleary, *supra* note 132, at 13; Kaye, *The Limits of the Preponderance of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation*, 1982 Am. B. Found. Research J. 487, 496-503. Applied to single-factor toxic tort cases, the long-term result of this rule is the payment by

B. *The Addition of the Attributable Risk Test to the Henle-Koch-Evans Postulates*

The Henle-Koch-Evans Postulates do not, by themselves, provide a complete legal standard because the determination of legal causation requires consideration of the degree of certainty required to meet the plaintiff's burden of proof. This deficiency can be remedied, however, by requiring in addition that the attributable risk for the factor at issue be greater than .50. Conceptually, the finder of fact must decide whether it is more likely than not that an individual plaintiff contracted a specific disease as a result of exposure to a factor for which the defendant is legally responsible. From an epidemiologic perspective, the question has two parts: (1) is the factor causally related to the disease (satisfaction of Henle-Koch-Evans Postulates), and (2) is the attributable risk greater than .50? If, in an exposed population, more than half the cases of a disease can be attributed to the exposure, and if the postulates are satisfied, then absent other information about a diseased individual, it is more likely than not that his or her illness was caused by the exposure.[143]

C. *Practical Application of the Evidentiary Test*

Consider a manufacturing plant that employs 1000 production workers. At some work stations widget grinders emit widget dust. Studies of people exposed to this type of dust for ten or more years at concentrations higher than 100 dust particles per cubic centimeter have indicated a relative risk of 2.5 (compared to non-exposed per-

---

defendants, taken collectively, of the total cost of the injuries they have caused to plaintiffs, taken collectively. The rule may break down in multi-factor cases, or in cases in which a defendant has very probably caused many, but not all, occurrences of a given disease in a relatively large population. In the latter situation either under-compensation or over-compensation of the plaintiffs, as a group, may result. These issues are discussed in the context of proportional liability, *infra* pt. V(B).

143. In using the Henle-Koch-Evans Postulates as constrained by attributable risk, great care must be taken in defining the exposure and the exposed population. In some instances, the focus should be on the total exposure above a certain level; in other cases the extent of exposure at any given time may be more important. The population of interest should be limited to individuals exposed at or beyond the level or extent at issue. For example, if the defined population included all steelworkers, it would be difficult to make inferences about the effects of prolonged high exposure to blast furnace fumes. New steelworkers and those who worked in rolling mills would not have suffered the same level of exposure as long-time blast furnace workers. To appreciate fully the problems that can be caused by improperly defining a population, consider a numerical example. Suppose that 10 of 50 blast furnace workers have a lung disease, that 100 of 1950 other steel workers have the same disease, and that 50 of 1000 non-steelworkers have it. Comparing the blast furnace workers to the general population yields a relative risk of 4, but if all steelworkers are considered, the relative risk drops to 1.1.

sons) for megabonkoma, a deadly (though fictional) form of lung cancer. If one of the widget workers contracts this terrible disease, could he establish through an epidemiologic study that it more likely than not resulted from widget dust exposure at the factory? Answering this question requires determining if the study results satisfy the Henle-Koch-Evans Postulates, and if the worker in question was exposed to widget dust for a long enough period and at a high enough concentration.

To test evidence against the Henle-Koch-Evans Postulates one must consider a number of factors. For example, breathing dust is more likely to cause a lung disease, such as megabonkoma, than a bone disease. This would support the inference of a causal connection. Studies that indicate a correlation between megabonkoma in rats and exposure to widget dust would tend to confirm human data and would further support the inference. Such biological information, together with a sufficiently large population sample, an absence of serious biases and a consistent and verified relative risk of 2.5 would probably support the inference that widget dust causes some cases of megabonkoma. The widget worker, however, would still have to establish both exposure and a sufficiently high attributable risk.

If the worker in question had held his job for over ten years, and had worked in a part of the factory where widget dust exceeded 100 particles per cubic centimeter, exposure would be quite clear, and the attributable risk of .60 would easily satisfy the more-likely-than-not test.[144] For situations in which sufficient exposure is certain, any relative risk greater than 2 would lead to an attributable risk of more than .50.[145] More typical, however, is the situation in which exposure is questionable. Perhaps the worker performed a number of tasks at various locations in the plant or used different machines that emitted varying amounts of dust. Under these circumstances, one could estimate the probability that exposure exceeded the level in the study. If only sixty percent of the 1000 workers were heavily exposed, the attributable risk would drop to .47,[146] even with a relative risk of 2.5. This evidence would fail the more-likely-than-not test and would not support a plaintiff's verdict.

---

144. $$\frac{1.0\ (2.5 - 1)}{1.0\ (2.5 - 1) + 1} = .60$$
See *supra* note 123 and accompanying text.

145. If the proportion of the populations exposed is 1.0, as in *supra* note 144, then:
$$\frac{1.0\ (2 + z - 1)}{1.0\ (2 + z - 1) + 1} = \frac{(1 + z)}{(2 + z)}$$
which is greater than 0.50 for any positive z.

146. $$\frac{.6\ (2.5 - 1)}{.6\ (2.5 - 1) + 1} = .47$$
Note that this example is somewhat oversimplified. It assumes that at any exposure less than 10 years and 100 particles per cubic centimeter the relative risk is 1.0.

A worker who contracted megabonkoma after high exposure for less than ten years might still be able to establish causation if he could produce evidence that the total amount of dust inhaled was an adequate measure of exposure. A person exposed at a relatively low level for more than ten years could make a similar argument. In no case, however, can evidence suffice to establish a causal link if it does not include at least reasonable estimates of exposure levels and durations, and data that reasonably indicate a relative risk greater than 2.[147]

IV. Precedents and Requirements for the Introduction of Epidemiologic Evidence

A party seeking to introduce scientific evidence faces two general requirements: The methods used to obtain data and to draw inferences therefrom must be legally acceptable, and the witnesses through whom the evidence is introduced must be suitably qualified.[148] Precedent supports not only admitting epidemiologic proof into evidence,[149] but also requiring that such proof be produced by a toxic tort plaintiff. Precedent also supports a rule requiring that a medical expert be qualified as an epidemiologist before testimony on causation is admitted in a toxic tort case.

A. *Precedents for Admitting Epidemiologic Proof into Evidence*

In cases involving diseases caused by viruses or bacteria, courts have generally accepted epidemiologic evidence with little difficulty,[150] and

147. The foregoing discussion leaves open many questions about the detailed application of the Henle-Koch-Evans Postulates, and about what constitutes a reasonable indication of relative risk. In actual cases, expert witnesses would probe the many complications and subtleties that have been omitted. At least one court has recognized that there is "room for responsible epidemiologists to differ significantly on many of the key choices and assumptions to be made in analyzing [a] causal relationship." O'Gara v. United States, 560 F. Supp. 786, 789 (E.D. Pa. 1983). To be sufficient, however, the testimony of experts should fall within the proposed framework.

148. *See* 3 J. Weinstein & P. Berger, Weinstein's Evidence §§ 702[01]-[04] (1982).

149. For an excellent bibliography and discussion of the admissibility and use of scientific evidence, see *Symposium on Science and the Rules of Evidence*, 99 F.R.D. 187 (1983). *See generally* Gianelli, *The Admissibility of Novel Scientific Evidence:* Frye v. United States, *a Half-Century Later*, 80 Colum. L. Rev. 1197, 1235-45 (1980) (discussion of the standards used to determine admissibility); Korn, *Law, Fact, and Science in the Courts*, 66 Colum. L. Rev. 1080, 1108-1113 (1966) (discussion of the process through which courts incorporate scientific principles and discoveries); McCormick, *Scientific Evidence: Defining a New Approach to Admissibility*, 67 Iowa L. Rev. 879, 882-83 (1982) (same).

150. *See, e.g.*, Kehm v. Procter & Gamble Mfg. Co., 724 F.2d 613, 617-20 (8th Cir. 1983) (toxic shock syndrome case in which court admitted into evidence epidemiologic reports from the Center for Disease Control); Wolf v. Procter & Gamble Co., 555 F. Supp. 613, 624-26 (D.N.J. 1982) (same); Travelers Ins. Co. v. Donovan,

there exists no rationale for treating such evidence differently in toxic tort cases. In fact, even in some toxic tort cases, courts have alluded to the concept of comparing incidence rates.[151] Some commentators have objected to this approach because the evidence is not specific to the plaintiff,[152] but they ignore the fact that even " '[p]articularistic' evidence offers nothing more than a basis for conclusions about a perceived balance of probabilities."[153] Other commentators have lamented that courts tend not to accept epidemiology,[154] but the basis for this assertion is unclear. In fact, good epidemiologic evidence is not only accepted by courts; in at least one case, it has been required.[155]

B. *Precedents for Incorporation of Epidemiologic Postulates into an Evidentiary Standard*

A number of precedents amply support an evidentiary standard incorporating scientific principles and requiring that evidence conform to them. Some courts have even measured evidence against the Ewing Postulates,[156] despite serious questions about their validity in legal proceedings and problems in applying them objectively. The postulates of epidemiology are far better established than Ewing's, and should be more readily used as the basis for a standard against which to test the sufficiency of evidence. Insofar as epidemiology involves statistics, decisions in a number of cases, not all involving toxic torts, have demonstrated the ability of courts to judge intelligently the validity of statistical inferences.[157]

1. Discrimination Cases

In discrimination cases, which often hinge on the statistical significance of the difference between the composition of a population and

---

125 F. Supp. 261, 262 (D.D.C. 1954) (tuberculosis case in which workers' compensation claimant was awarded recovery based on increased a priori risk), *aff'd*, 221 F.2d 886 (D.C. Cir. 1955); Sacred Heart Med. Center v. Carrado, 92 Wash. 2d 631, 637, 600 P.2d 1015, 1019 (1979) (hepatitis case in which recovery was allowed based on plaintiff's elevated a priori risk of contracting the disease).

151. See *supra* note 58 and accompanying text.

152. Dickson, *supra* note 57, at 799-808; Dore, *supra* note 57, at 431.

153. Rosenberg, *supra* note 5, at 870.

154. See *supra* note 57.

155. Heyman v. United States, 506 F. Supp. 1145, 1149 (S.D. Fla. 1981). See *infra* notes 180-82 and accompanying text.

156. *See, e.g.*, Stordahl v. Rush Implement Co., 148 Mont. 13, 19-20, 417 P.2d 95, 99 (1966); Sikora v. Apex Beverage Corp., 282 A.D. 193, 196, 122 N.Y.S.2d 64, 66 (1953), *aff'd*, 306 N.Y. 917, 119 N.E.2d 601 (1954); Dennison v. Wing, 279 A.D. 494, 496-97, 110 N.Y.S.2d 811, 813-14 (1952).

157. *See* Rosenberg, *supra* note 5, at 870-71.

the composition of a work force, jury panel or the like, courts have shown great understanding of the value of testing hypotheses against data. The Supreme Court, in two 1977 discrimination cases,[158] explicitly approved the type of significance testing used in the statistical part of an epidemiologic study. A third case decided that year involved similar though less explicit reasoning.[159] *Castaneda v. Partida*[160] dealt with grand jury selection practices in Hidalgo County, Texas. Although the population was approximately eighty percent Hispanic, grand jury participation over a ten-year period averaged only thirty-nine percent Spanish surnamed, with the highest annual figure just over fifty percent.[161] The chance of such disproportionate representation was extremely low, assuming no discrimination. The Court, therefore, rejected this null hypothesis[162] and held that, absent rebuttal evidence, the alternative hypothesis of discrimination should be accepted.[163] *Hazelwood School District v. United States*[164] and *International Brotherhood of Teamsters v. United States*[165] involved discriminatory hiring practices alleged to be in violation of Title VII of the Civil Rights Act of 1964.[166] In *Teamsters*, the Court explicitly approved the use of statistics to establish a prima facie case of discrimination, but did not delve into details of methodology.[167] In *Hazelwood*, however, it endorsed the more rigorous statistical approach used in *Castaneda*.[168]

These and subsequent cases[169] clearly establish the ability of courts to understand and use classical hypothesis testing techniques. They do not, however, address the basic issue of tort causation. A statistically significant difference in a discrimination case shifts the burden of

158. Hazelwood School Dist. v. United States, 433 U.S. 299 (1977); Castaneda v. Partida, 430 U.S. 482 (1977).

159. International Bhd. of Teamsters v. United States, 431 U.S. 324 (1977).

160. 430 U.S. 482 (1977).

161. *Id.* at 486-87 & n.7.

162. *See id.* at 494 & n.13. See *supra* note 104.

163. 430 U.S. at 496 n.17. The Court noted that the likelihood that random selection would produce the jury panels actually selected in Hidalgo County was less than 1 in 10^{140}.

164. 433 U.S. 299 (1977).

165. 431 U.S. 324 (1977).

166. *Hazelwood*, 433 U.S. at 301; *Teamsters*, 431 U.S. at 328.

167. 431 U.S. at 339.

168. 433 U.S. at 308 n.14, 311 n.17.

169. *See, e.g.*, Plemer v. Parsons-Gilbane, 713 F.2d 1127, 1137 (5th Cir. 1983); Harris v. Birmingham Bd. of Educ., 712 F.2d 1377, 1383 (11th Cir. 1983); Chisholm v. United States Postal Serv., 665 F.2d 482, 494-95 (4th Cir. 1981); EEOC v. United Virginia Bank, 615 F.2d 147, 149-54 (4th Cir. 1980).

proof, and absent rebuttal evidence, establishes the plaintiff's allegations as facts. In a toxic tort case, the difference not only must be statistically significant, but also must be sufficiently large to make it more likely than not that the individual plaintiff's injury resulted from the defendant's substance. The standard proposed in this Article would also require consistency with the Henle-Koch-Evans Postulates.

The following example illustrates the difference between discrimination and toxic tort cases. Evidence that a company employs a workforce that is only fifteen percent black from a population that is twenty percent black might conclusively prove discrimination. A twenty percent disease rate in a population exposed to a chemical, however, would not prove tort causation in an individual case if the unexposed population experienced a fifteen percent rate. The maximum attributable risk would be only twenty-five percent.[170] Moreover, even if the exposed population had a disease rate of forty-five percent, and the attributable risk were as high as sixty-seven percent,[171] the Henle-Koch-Evans Postulates would still have to be satisfied.

2. Identity Cases

Criminal law is another area in which courts have examined statistical evidence. Proving the identity of a criminal often involves the use of circumstantial evidence indicating that certain events would be very unlikely to occur by coincidence. Attempts to quantify this mode of proof through statistics, however, have generally foundered. The best known example is *People v. Collins*,[172] in which a white woman and a black man in a yellow car committed an assault and robbery. The defendants fit this description, and the prosecution introduced evidence that the probability of these factors occurring together by coincidence was extremely slight. The jury found the defendants guilty, but the Supreme Court of California overturned the conviction because the method used to compute the probability of coincidence was flawed, and because the unlikelihood of coincidence did not establish the probability that the accused couple was guilty.[173] In a large population, even a rare combination could be expected to occur more than just once. Therefore, without resorting to a controversial

170. $$\frac{1\ (20/15\text{-}1)}{1\ (20/15\text{-}1) + 1} = .25$$
See *supra* note 123 and accompanying text.

171. $$\frac{1\ (45/15\text{-}1)}{1\ (45/15\text{-}1) + 1} = .67$$
See *supra* note 123 and accompanying text.

172. 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968).

173. *Id.* at 327-31, 438 P.2d at 38-40, 66 Cal. Rptr. at 502-05.

technique known as Bayesian analysis,[174] one cannot directly ascribe a probability to the hypothesis of guilt.[175]

The *Collins* problem limits all hypothesis testing. It is because statistics in themselves do not determine probability[176] that the non-statistical postulates of epidemiology are so extremely important. To reach a scientifically sound opinion that a causal link more likely than not exists, one must integrate other information. The overall epidemiologic approach and the need for a substantive standard are both illustrated by the swine flu cases, which constitute the best judicial use of epidemiology to date.

3. Swine Flu Cases

In 1976, fear of an impending influenza epidemic prompted rapid implementation of a swine flu inoculation program[177] before final testing of the vaccine could be completed. As a result, no drug company would manufacture the vaccine until the federal government agreed to assume all liability.[178] Thus, the swine flu cases were tried under the Federal Tort Claims Act[179] before federal district judges and without a jury, and many of the opinions did not reach the issue of

---

174. *See* Finkelstein & Fairley, *A Bayesian Approach to Identificiation Evidence*, 83 Harv. L. Rev. 489, 498-501 (1970).

175. *Id. But see* Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 Harv. L. Rev. 1329, 1365-76 (1971) (rejecting the use of Bayesian analysis).

176. For paternity cases, some courts, *e.g.*, Cramer v. Morrison, 88 Cal. App. 3d 873, 884-85, 153 Cal. Rptr. 865, 871-72 (1979); Malvasi v. Malvasi, 167 N.J. Super. 513, 515-16, 401 A.2d 279, 280 (1979); *see, e.g.*, Lascaris v. Laredo, 100 Misc. 2d 220, 221-23, 417 N.Y.S.2d 665, 666-67 (1979), legislatures, *e.g.* Ariz Rev. Stat. Ann. § 12-847(c) (West 1982), and even the ABA, Abbott, *Joint AMA-ABA Guidelines: Present Status of Serologic Testing in Problems of Disputed Parentage*, 10 Fam. L.Q. 247, 257 (1976), have embraced statistics without giving adequate attention to this problem. The accuracy of modern blood-typing techniques permits the exclusion of at least 90% of falsely identified men. That is, rejection of the null hypothesis of non-paternity is wrong only about 10% of the time. This does not, however, establish the probability that the alternative hypothesis is true. It only means that the null hypothesis does not conform well to the data. What then is the probability that a man not excluded is in fact the father of the child? The answer cannot be derived solely from the test results. Important assumptions about the number of possible putative fathers must be made. For a discussion of the misuse of the new testing techniques, see Ellman & Kaye, *Probabilities and Proof: Can HLA and Blood Group Testing Prove Paternity?*, 54 N.Y.U. L. Rev. 1131 (1979).

177. For a discussion of the history of the swine flu program, see *In re* Swine Flu Immunization Prods. Liab. Litig., 533 F. Supp. 567, 571-72 (D. Colo. 1980).

178. *See id.* at 572. The government agreed to assume liability because otherwise the manufacturers would be subject to strict products liability claims for any defects in the manufacture of the vaccine. *Id.*

179. 28 U.S.C. §§ 1346(b), 2401(b), 2671-2680 (1976 & Supp. V 1981).

legal sufficiency. They did, however, discuss in detail the judicial evaluation of the evidence involved.

Epidemiologic analysis figured decisively in most of the swine flu cases. For example, the court in *Heyman v. United States*[180] rejected the plaintiff's claim because she attempted to prove her case without epidemiologic evidence. The court found that clinicians generally cannot determine "whether a relationship exists between an illness and a preceding event such as a vaccination,"[181] and held that "without at least some reference to epidemiological studies, [the] plaintiff's position that her illness was caused by the swine flu shot amounts to nothing more than speculation."[182]

Central to the swine flu litigation was an epidemiologic study that indicated a relative risk of greater than 2 for Guillain-Barre Syndrome (GBS) up to ten weeks after swine flu inoculation.[183] If the exposed (vaccinated) population is perfectly defined, a relative risk of 2 corresponds to an attributable risk of 0.50.[184] The study induced the government to settle almost all cases in which the plaintiff contracted GBS within ten weeks of his swine flu shot.[185] Thus, the plaintiffs in virtually all of the reported cases either contracted GBS more than ten weeks after their inoculations, or contracted a disease other than GBS. One of these cases exemplifies the proper use of epidemiology; another shows the need for an epidemiologic evidentiary standard.

In *Cook v. United States*,[186] the plaintiff GBS victims experienced the onset of the disease approximately twelve weeks after their swine flu inoculations.[187] The district court disallowed their claims after a detailed and perceptive discussion of the use of epidemiology.[188] The judge discussed the connection between a relative risk of 2 and the more-likely-than-not standard, and also determined that the court should consider the two non-statistical factors of alternative explanations and biological credibility. [189]

---

180. 506 F. Supp. 1145 (S.D. Fla. 1981).
181. *Id.* at 1149.
182. *Id.*
183. Schonberger, Bregman, Sullivan-Bolyai, Keenlyside, Ziegler, Retailliau, Eddins & Bryan, *Gullian-Barre Syndrome Following Vaccination in the National Influenza Immunization Program, United States, 1976-1977*, 110 Am. J. Epidemiology 105, 112-13 (1979). The study also discussed attributable risk, *id.* at 111-13, but this was not used in any of the legal analyses.
184. $\frac{1\,(2\text{-}1)}{1\,(2\text{-}1)+1} = .50$
185. Hall & Silbergeld, *supra* note 5, at 446.
186. 545 F. Supp. 306 (N.D. Cal. 1982).
187. *Id.* at 307; *see* Padgett v. United States, 553 F. Supp. 794, 804 (W.D. Tex. 1982).
188. 545 F. Supp. at 315-16.
189. *Id.* at 314-15.

At the other extreme is *Sulesky v. United States*,[190] in which the plaintiff first exhibited signs of GBS more than three months after her injection.[191] She introduced epidemiologic testimony that conflicted with a government report that had previously been relied on in many cases.[192] This so confused the court that it turned to the testimony of treating and evaluating physicians, who apparently did not even discuss the disease's relative rate of occurrence.[193] Nonetheless, the court, relying on their testimony, held for the plaintiff.[194] Without a substantive standard for review, an appellate court faced with the *Cook* and *Sulesky* verdicts would have to uphold both, although only the first could be rationally explained. This Article's proposal would provide both trial and appellate courts with the required standard.

C. *Qualifications for Expert Witnesses Giving Testimony About Epidemiology*

No court has yet determined the qualifications necessary for a witness to offer expert testimony about epidemiology. In general, a witness need only have such "knowledge, skill, experience, training, or education" in the field at issue to "make it appear that his opinion . . . will probably aid the trier in his search for truth."[195] In *Jenkins v. United States*[196] it was held that a psychologist could, under some circumstances, give psychiatric testimony. The court cited an earlier case in which it had been held that "a general practitioner may testify concerning matters within a medical specialty if his education or experience, or both, involves demonstrable knowledge of the subject."[197]

Under the *Jenkins* test and the proposed standard, a medical doctor could testify about toxic tort causation only if he could demonstrate knowledge of epidemiology.[198] The preference often accorded treating physicians should not apply because a standard based on the drawing of inferences from populations does not require detailed knowledge of the plaintiff's individual case. Moreover, a medical degree would not

---

190. 545 F. Supp. 426 (S.D. W. Va. 1982).
191. *Id.* at 429.
192. *Id.* at 429-30.
193. *See id.* at 430-31.
194. *Id.* at 431.
195. Fed. R. Evid. § 702; *see* Jenkins v. United States, 307 F.2d 637, 643 (D.C. Cir. 1962) (quoting McCormick, Law of Evidence § 13 (1954)).
196. 307 F.2d 637 (D.C. Cir. 1962).
197. *Id.* at 643-44 (citing Sher v. DeHaven, 199 F.2d 777, 782 (D.C. Cir. 1952)).
198. Sufficient knowledge might be established in a number of ways, including coursework or membership in appropriate professional organizations. In Kubs v. United States, 537 F. Supp. 560 (E.D. Wis. 1982), a witness' testimony was rejected in part because the work on which it was based had never been subjected to peer review. *Id.* at 562.

necessarily be required because many epidemiologists do not have one. The best witness, of course, would be a medical doctor thoroughly trained in epidemiology, because the need to integrate biology, statistics and common sense to draw proper inferences requires as broad a background as possible.

V. "First Case," Under-Compensation and Over-Compensation Problems

A. *The "First Case" Problem*

The use of epidemiology to determine the legal sufficiency of evidence would eliminate much of the inconsistency and irrationality from judicial decisions in which the causation of a latent disease is at issue. It would, however, also make it difficult for victims to prove causation prior to the development of adequate data, and thus would create a special problem for early victims.[199] In response to this "first case" problem, some commentators have proposed scientifically questionable rules to ease the plaintiff's burden of proof.[200] These proposals would, in effect, remove all rational limits on liability, a step for which most proponents give no theoretical justification beyond an unfocused desire to compensate.[201]

The unlimited liability that would result from relaxed evidentiary standards is best illustrated by one proposal that would require a

---

199. Several of the swine flu cases involved claims that diseases such as polymyostis, Tabaczynski v. United States, 529 F. Supp. 156, 161 (E.D. Mich. 1981), *aff'd*, 711 F.2d 1059 (6th Cir. 1983), or arthritis, Gicas v. United States, 508 F. Supp. 217, 220 (E.D. Wis. 1981), were caused by swine flu inoculations. Most were rejected because a single isolated temporal coincidence is not sufficient evidence. One expert in *Tabaczynski* pointed out that a single case could not support statistical inferences. 529 F. Supp. at 62. *But see* Hasler v. United States, 517 F. Supp. 1262, 1271 (E.D. Mich. 1981) (onset of arthritis after inoculation was found not to be a coincidence), *rev'd*, 718 F.2d 202 (1983).

The modification of the statute of limitations in several states to allow a plaintiff to bring an action after a causal relationship is discovered is implicitly based on the recognition that scientifically establishing causation often requires time for the accumulation of data. *See* Stoleson v. United States, 629 F.2d 1265, 1269 (7th Cir. 1980).

200. *See, e.g.*, Hall & Silbergeld, *supra* note 5, at 442-43 (extrapolation of epidemiologic studies representing unusual subgroups in population to larger group to establish causation based on assumption that different species react similarly to different substances); *Tort Actions for Cancer*, *supra* note 5, at 855-59 (calling for government maintenance of a catalog of exposure levels at which particular carcinogenic substances will cause cancer; presumption of causation is created if plaintiff can show exposure above the threshold level).

201. *See* Burcat, *supra* note 5, at 857-59; Soble, *A Proposal for the Administrative Compensation of Victims of Toxic Substance Pollution: A Model Act*, 14 Harv. J. on Legis. 683, 768 (1977); *Precursor Symptoms*, *supra* note 2, at 194; *Environmental Risks*, *supra* note 5, at 587.

plaintiff to prove only exposure "significant enough to trigger disease."[202] According to some theories, significance could be found in very low exposures.[203] Thus adoption of the proposal could make nearly everyone potentially liable to countless people. Simply breathing releases traces of suspect organic compounds.[204]

Other proposals would use methods employed in making regulatory decisions to establish rebuttable presumptions of causation in tort actions.[205] Rebuttal, however, would require the same kind of studies needed to establish causation under the proposed standard. If the burden of proof were shifted to defendants in this way, the loss, in the absence of any information on causation, would be transferred to them.[206] Such a change would be at odds with recognized legal principles.

A clear distinction currently exists between the standard of proof used in regulation and the standard used in determining tort liability. Most legislation governing the regulation of potentially toxic substances requires far less convincing proof of harmfulness than would satisfy the more-likely-than-not test.[207] As a result, regulatory agencies have employed methods that would not meet the proposed test of evidentiary sufficiency. In particular, agencies have banned or limited certain substances on the basis of animal studies backed by little, if any, human data.[208] This type of analysis may be appropriate in protective regulation, but it does not satisfy the more-likely-than-not test[209] and should not, as some have argued, carry over to tort cases.[210]

---

202. Hall & Silbergeld, *supra* note 5, at 445.

203. S. Epstein, The Politics of Cancer 3 (1978).

204. *See* ABA Section of Science & Technology, *Law, Science and Technology in Health Risk Regulation II*, 22 Jurimetrics J. 380, 381 (1982) (statement of Dr. Leon Golberg).

205. See *supra* note 5.

206. *See* Robinson, *supra* note 5, at 729 (placement of the burden of proof is dispositive of factual issue of causation); Rosenberg, *supra* note 5, at 866 n.65 ("shifting the burden would simply replace one bias with another").

207. *See* Reserve Mining Co. v. EPA, 514 F.2d 492, 520 (8th Cir. 1975) (reasonable medical concern for public health suffices to sustain agency action), *modified*, 529 F.2d 181 (1976); Environmental Defense Fund v. EPA, 510 F.2d 1292, 1298 (D.C. Cir. 1975) (the standard of "substantial evidence" means something less than the weight of the evidence); *see also* Maines, *Offensive Collateral Estoppel in Mass Tort or Products Liability Cases: The Potential for Corporate Catastrophe from Prior Administrative Proceedings*, 35 Admin. L. Rev. 327, 329-30 (1983) (lesser standard of proof in administrative hearings). *But see* Industrial Union Dep't v. American Petroleum Inst., 448 U.S. 607, 653 (1980) (more-likely- than-not test).

208. Environmental Defense Fund V. EPA, 510 F.2d 1292, 1299 (D.C. Cir. 1975).

209. Latin, *The "Significance" of Toxic Health Risks: An Essay on Legal Decision-making Under Uncertainty*, 10 Ecology L.Q. 339, 377-80 (1982).

210. *Id.*

The pitfalls of making conclusory legal leaps from mouse to man prevent rational extrapolation even in apparently extreme cases. Dioxin, for example, is a potent human toxin that may also be a carcinogen. In animals, its carcinogenic potency exeeds that of aflatoxin B, known as perhaps the most potent human carcinogen.[211] Its toxic effects, however, vary by a factor of 5000 in comparisons between tests using guinea pigs and hamsters.[212] If dose-response information for guinea pigs does not apply to another species of rodent, animal data are obviously not a reliable basis for making quantitative conclusions about exposed humans.[213] For regulatory purposes the existing evidence about dioxin may support the most stringent of limitations, but statements about likelihood in tort cases require more. Even within the regulatory context courts have recognized that human epidemiologic data should be given more weight than the results of animal testing. In *Dow Chemical Co. v. Blum*,[214] an epidemiologic study, albeit weak, sufficed to sustain an EPA order banning certain herbicides,[215] while in *Gulf South Insulation v. Consumer Product Safety Commission*,[216] the failure to consider epidemiologic evidence resulted in the reversal of a ban on the use of urea-formaldehyde foam insulation.[217]

The use of regulatory or other lesser standards in tort actions has been advocated by at least one proponent as necessary to achieve the tort system's goals of compensation, deterrence and retributive justice.[218] Relaxing evidentiary standards, however, would serve only the goal of compensation[219] and would unjustifiably single out the victims of certain diseases for special treatment. If society's only concern is compensation, why should lung cancer victims receive money when cystic fibrosis or multiple sclerosis victims do not? Even more to the point, why should a lung cancer victim who can demonstrate an exposure speculatively related to the disease receive compensation when other victims do not? If compensation is the only goal, it is best uncoupled from tort liability.

In addition to creating a crazy-quilt pattern of payments to disease victims, focusing only on compensation would seriously impair the

211. Friedman & Weckesser, *Dioxin and Resource Recovery*, Envtl. F., Sept. 1983, at 44, 46.

212. Rawls, *Dioxin's Human Toxicity is Most Difficult Problem*, Chem. & Eng'g News, June 6, 1983, at 37; *see* Mays, *Dioxin: Deadly or Deceptive?*, Envtl. F., Feb. 1984, at 13, 14.

213. *See* Mays, *supra* note 212, at 13-14.

214. 469 F. Supp. 892 (E.D. Mich. 1979).

215. *Id.* at 907.

216. 701 F.2d 1137 (5th Cir. 1983)

217. *See id.* at 1146.

218. *Environmental Risks*, *supra* note 5, at 575.

219. *Id.*

other goals of the tort system. Retribution unrelated to fault and causation is meaningless. As to deterrence, defendants would have little incentive to alter their conduct because they would be held liable to many victims even when they did not in fact cause their injuries. Most potential targets of toxic tort litigation are industrial concerns, but cancer and other latent diseases would continue to occur even if they completely ceased production. Far fewer cancers are tied to specific substances or activities than many have assumed.[220]

---

220. It is generally estimated that 60-90% of all cancers are linked in some way to the environment. Sixth Annual Report of the Council on Environmental Quality 33 (1975). This does not mean, however, that prevention of 60-90% is practicable. *See* Doll & Peto, *supra* note 55, at 1205-07; Higginson & Muir, *Environmental Carcinogenesis: Misconceptions and Limitations to Cancer Control*, 63 J. Nat'l Cancer Inst. 1291, 1296 (1979). Nor does it mean that man-made chemicals cause most cases of the disease, as some have concluded. *See Tort Actions for Cancer*, *supra* note 5, at 840-41. Even were this true, it would not justify relaxing evidentiary standards to facilitate almost universal recovery against chemical manufacturers when specific chemicals are not implicated.

Despite the complexity of environmental carcinogenesis, advocates of reducing the plaintiff's burden of proof have based their proposals in part on estimates that 20-40% of all cancers result from workplace exposures. The source of this estimate and its infiltration into legal commentary is an interesting story in itself. In 1978 a group of scientists from several federal agencies put together a report that contained the 20-40% figures. Occupational Factors, *supra* note 93, at 22. This report was widely criticized, and at least two of the authors later conceded that they had "relied on some assumptions about data that have been shown subsequently to be incorrect." Davis, Bridbord & Schneiderman, *Estimating Cancer Causes: Problems in Methodology, Production, and Trends*, 9 Banbury Rep. 285, 308 (1981). Nonetheless the estimates were defended by others, including Dr. Samuel Epstein, one of the most vocal critics of the American industrial establishment's use and control of suspect substances. Epstein & Swartz, *Fallacies of Lifestyle Cancer Theories*, 289 Nature 127 (1981). Dr. Epstein's book, The Politics of Cancer, *supra* note 203, formed the basis for many of the assertions on which one of the proposals for shifting the burden of proof was based. *See Tort Actions for Cancer*, *supra* note 5, at 848-50 (analysis focusing on the overall relationship of cancers to chemicals but implicitly incorporating Dr. Epstein's use of the high occupational estimates); Note, *Occupationally Induced Cancer Susceptibility: Regulating the Risk*, 96 Harv. L. Rev. 697, 697 n.3 (1983) (dealing with regulation rather than tort law, citing the original 20-40% estimate). Thus, legal commentators persist in propagating scientific overstatement.

To obtain such high figures, one must unrealistically assume that all workers are exposed to potential carcinogens at the highest reported rates. *See* Doll & Peto, *supra* note 55, at 1240-41. One realistic epidemiologic analysis of the occupational cancer issue indicates that 4% is a far more appropriate estimate. *Id.* at 1245. Other reasonable estimates range from 1% to 10%. Wynder & Gori, *supra* note 55, at 830. It has not been determined what portion of this 1-10% receives legal compensation, but it is clear that for occupationally-caused cancer the potential for tort system dysfunction is, at worst, far less than the actual dysfunction assumed by supporters of relaxed standards. Furthermore, environmental exposures are generally much less concentrated than those experienced in the workplace, indicating that the overall dysfunction is exaggerated as well.

Thus, the first case problem does not warrant changes in tort law principles. It does, however, still require that the difficulty of collecting sufficient data to satisfy an evidentiary standard derived from epidemiology be addressed. There are legal and institutional reforms that would reduce this burden without compromising principles or creating unlimited liability. One of the most irrational barriers to recovery results when the statute of limitations precludes a claim because a disease manifests itself too long after exposure,[221] or when scientific knowledge linking exposure and disease comes too late after manifestation. As is already the law in many states,[222] the statutory period should commence with a plaintiff's illness, if the causal link is known at that time, or when causation becomes reasonably apparent.[223]

Another appropriate legal reform would be the adoption of procedural changes at the state level that would facilitate joint collection of evidence by plaintiffs. Within the federal court system, consolidation of cases for purposes of discovery has already proven useful in mass personal injury cases.[224] So, too, has the federal class action device.[225]

---

221. *See, e.g.*, Steinhardt v. Johns-Manville Corp., 54 N.Y.2d 1008, 430 N.E.2d 1297, 446 N.Y.S.2d 244 (1981). In *Steinhardt*, the New York Court of Appeals held that an action for disease resulting from occupational exposure to asbestos was barred by the statute of limitations because it was commenced more than four years after the plaintiff's last employment-related exposure. *Id.* at 1010, 430 N.E.2d at 1298-99, 446 N.Y.S.2d at 245-46.

222. Annot., 1 A.L.R. 4th 117, 127-34 (1980).

223. *See, e.g.*, Large v. Bucyrus-Erie Co., 707 F.2d 94, 96-97 (4th Cir. 1983); Grabowski v. Turner & Newell, 516 F. Supp. 114, 118-20 (E.D. Pa.), *aff'd*, 651 F.2d 908 (3d Cir. 1980) (per curiam). Locke v. Johns-Manville Corp., 221 Va. 951, 958-59, 275 S.E.2d 900, 905 (1981). For a discussion of this rule as applied in federal courts, see Davis v. United States, 642 F.2d 328, 331 (9th Cir. 1981), a case involving polio vaccine. The *Davis* court refused to delay tolling of the statute until negligence as well as causation was discovered.

224. *E.g.*, *In re* Swine Flu Immunization Prods. Liab. Litig., 446 F. Supp. 244, 246 (J.P.M.D.L. 1978) (per curiam), *vacated*, 687 F.2d 14 (1982); *In re* A.H. Robins Co., "Dalkon Shield" IUD Prods. Liab. Litig., 419 F. Supp. 710, 712 (J.P.M.D.L. 1976) (per curiam); *In re* A.H. Robins Co., "Dalkon Shield" IUD Prods. Liab. Litig., 406 F. Supp. 540, 542 (J.P.M.D.L. 1975) (per curiam); *see* Note, *The Judicial Panel and the Conduct of Multidistrict Litigation*, 87 Harv. L. Rev. 1001, 1002-09 (1974).

225. *E.g.*, *In re* Three Mile Island Litig., 87 F.R.D. 433, 442 (M.D. Pa. 1980); Payton v. Abbott Labs., 83 F.R.D. 382, 387-88 (D. Mass. 1979). Other cases have denied class certification. *E.g.*, *In re* Northern Dist. of Cal., Dalkon Shield IUD Prods. Liab. Litig., 693 F.2d 847, 850-51 (9th Cir. 1982), *cert. denied*, 103 S. Ct. 817 (1983); *In re* Federal Skywalk Cases, 680 F.2d 1175, 1182-83 (8th Cir.), *cert. denied*, 103 S. Ct. 342 (1982); Ryan v. Eli Lilly & Co., 84 F.R.D. 230, 234 (D.S.C. 1979); *see* Seltzer, *Punitive Damages in Mass Tort Litigation: Addressing the Problems of Fairness, Efficiency and Control*, 52 Fordham L. Rev. 37, 69-71 (1983). For a good bibliography on class actions, see McGovern, *Management of Multiparty Toxic Tort Litigation: Case Law and Trends Affecting Case Management*, 19 Forum 1, 9 n.18 (1983).

Not all states permit such combined efforts, and to the extent that they do not, more liberal rules should be adopted.[226]

Of course, scientific research is not optimally conducted with the primary aim of preparing data for litigation. For the long term, a coordinated research effort is required, an effort which certain institutional changes would promote. Increased funding for governmental agencies that collect and analyze epidemiologic data would be a first step, but government agencies should not do all the work.[227] The establishment of a fund for research at universities or by other relatively disinterested private sector individuals or groups would diversify the information gathering effort. This fund could be provided at least in part by interested industries. Few institutional mechanisms for such participation now exist, but a number of changes are possible. These range from direct payments by industry for research, made perhaps in conjunction with labor unions, to the establishment of an umbrella organization to distribute money paid according to some form of cost allocation system.[228]

---

226. 301(e) Study, *supra* note 3, at 257.

227. *See* Letter of Lilienfeld & Lilienfeld to the editors of Science, 198 Science 250-53 (Oct. 21, 1977) (suggesting a sort of Brookings Institution for science).

228. Research will not, of course, solve all causation problems, but even initial studies may have legal uses. In the case of some occupational diseases, it may be possible to establish a relationship between working in a particular industry and the incidence of the diseases, though great care must be exercised in interpreting the data. In one Finnish study of how cancer incidence varied by occupation, almost all of the differences were found to be associated with different cigarette consumption habits of the people who tended to go into the occupations under investigation. Pukkala, Teppo, Hakulinen & Rimpela, *Occupation and Smoking as Risk Determinants of Lung Cancer*, 12 Int'l J. Epidemiology 290, 293-95 (1983). For an example of how the performance of certain tasks within an occupation may be implicated, see American Iron & Steel Inst. v. OSHA, 577 F.2d 825, 832 (3d Cir. 1978) (relationship between lung cancer and exposure to coke oven emissions), *cert. dismissed*, 448 U.S. 917 (1980).

Evidence derived from occupational studies can facilitate proof within the workers' compensation context even without identifying a particular substance, and it can assist in further pinpointing the cause. The rubber industry provides an excellent example of how research studies can be done through industry-university cooperation without government funding or coercion. The 1970 union contract with the rubber industry provided for a comprehensive occupational research program. The Schools of Public Health at Harvard University and the University of North Carolina at Chapel Hill contracted with both the United Rubber Workers and the major U.S. rubber companies to do the work. They found, among other things, that certain cancers were more common in rubber workers than in the general population, though overall, the excess mortality for all cancers was minimal. *See* McMichael, Andjelkovic & Tyroler, *Cancer Mortality Among Rubber Workers: An Epidemiologic Study*, 271 Annals of the N.Y. Acad. of Science 125, 136 (1976); Monson & Fine, *Cancer Mortality and Morbidity Among Rubber Workers*, 61 J. Nat'l Cancer Inst. 1047 (1978). Similar industry-wide efforts should be encouraged in the future.

In addition to increasing and improving research, steps to insure better investigation of suspect cases of certain diseases could also be taken. The interests of society have already led to statutory requirements that certain deaths of unclear origin automatically fall under the jurisdiction of a coroner. These requirements greatly aid in collecting the information necessary to determine or prove if a crime has been committed. Similarly, to insure that occurrences of a disease possibly related to a toxic tort are properly and adequately documented, they should, by statute, come within the coroner's or some other health officer's jurisdiction.[229] This would not only make it easier to determine causation in the particular case under investigation, but would also generate data for general use in drawing epidemiologic inferences.

B. *Under-Compensation and Over-Compensation*

In some circumstances an epidemiologic standard would cause a radical shift from non-compensation to over-compensation. Until sufficient evidence of causation is developed, all plaintiffs would lose; afterwards, assuming there exists adequate proof of exposure and other necessary elements of the legal theory being pursued, all would likely win.[230] To avoid this dichotomy, and to allow some recovery when the burden of proof is not met, a few commentators have suggested proportional recovery.[231]

Under the proportional approach, if thirty percent of the a priori risk of a disease were attributable to a defendant, the plaintiff would recover thirty percent of his damages from that defendant. Likewise,

229. Autopsy of *all* deaths should be encouraged. Autopsy, which often reveals false diagnoses, is now on the decline in the United States, a trend that should be reversed. *See* Lundberg, *Autopsies as the Doctor's—and Patient's—Best Friend*, J. A.M.A., September 2, 1983, *reprinted in* Baltimore Sun, Oct. 30, 1983, at K5.

230. Whether collateral estoppel on the issue of causation would be applied is an open question. The issue has been considered in the context of the asbestos litigation. *See* Baldwin, *Asbestos Litigation and Collateral Estoppel*, 17 Forum 772, 781-83 (1982); Comment, *An Examination of Recurring Issues in Asbestos Litigation*, 46 Alb. L. Rev. 1307, 1330-31 (1982); Note, *Applying Offensive Collateral Estoppel to Asbestos Cases: A Viable Alternative*, 16 Suffolk U. L. Rev. 687, 702-06 (1982). Even without collateral estoppel, however, once sufficient evidence is collected for one case, it will probably be available for use in other cases. *See* McGovern, *supra* note 225, at 8. This militates against hasty expansion of the scope of collateral estoppel. *See generally* Maines, *supra* note 207 (discussing problems that might be created by expansion of the scope of collateral estoppel based on administrative findings).

231. *See, e.g.*, Estep, *supra* note 57, at 281-86; Rizzo & Arnold, *supra* note 5, at 1407-13; Robinson, *supra* note 5, at 743-49. One recent proposal for proportional recovery would combine the concept with rebuttable presumptions. *Environmental Risks*, *supra* note 5, at 614-15. For the reasons discussed *supra* notes 135-42, this proposal is highly questionable.

if the risk were seventy percent, recovery would be limited to that percentage of the damages. Such verdicts are not possible today.[232] Although some courts have apportioned liability among several defendants when the harmfulness of the substance involved was not at issue,[233] no such award has ever been based on the probability of harmfulness.[234]

Proportional recovery would shift the focus of legal analysis from the individual case to the tortfeasor who has caused many, but not all, injuries. It would allow some plaintiffs to recover who, given perfect information, would not. It would also mean something less than complete recovery for those who would receive full compensation under the traditional rules. The tortfeasor would pay the full cost of the damage it had caused, but not necessarily to the parties it actually injured. One commentator has described this result as being "actuarially fair,"[235] and another has justified it on the ground that the law should prefer inexact justice to manifest injustice.[236]

Whether proportional recovery would in fact be more just than the present all-or-nothing rule remains an open question. Adoption of the theory would not dramatically reduce the difficulties faced by toxic tort plaintiffs. Its rational implementation would require epidemiologic evidence similar to that required to satisfy the proposed standard. Attributable risks of less than fifty percent would not preclude recovery, but data to support reasonable estimates of attributable risk would still be required.

The net effect of proportional recovery would depend on factors that require further investigation and research. The feasibility of detecting small attributable risks must be determined. It must also be determined whether more cases involve attributable risks above fifty percent or below fifty percent. If relatively small impacts on incidence rates defy detection, a theory intended to assist plaintiffs who cannot

---

232. Apportionment today depends primarily on the nature of the plaintiff's injury, rather than on the conduct of defendants. *See* W. Prosser, *supra* note 11, § 52, at 314.

233. A good example is Sindell v. Abbott Labs., 26 Cal. 3d 588, 607 P.2d 924, 163 Cal. Rptr. 132, *cert. denied*, 449 U.S. 912 (1980), a DES case in which the court used a "market share" theory of liability. *Id.* at 611-13, 607 P.2d at 937, 163 Cal. Rptr. at 145; *see* Comment, *DES and a Proposed Theory of Enterprise Liability*, 46 Fordham L. Rev. 963, 995-1000 (1978). *But see* Sheffield v. Eli Lilly & Co., 144 Cal. App. 3d 583, 592-99, 192 Cal. Rptr. 870, 875-80 (1983) (refusing to apply the market share theory).

234. One court, however, by denying a defendant's motion for summary judgment, implied that the *Sindell* theory might apply to the issue of harmfulness as well as to identity of the party responsible for exposing the plaintiff. Pereira v. Dow Chem. Co., 129 Cal. App. 3d 865, 872-73, 181 Cal. Rptr. 364, 368 (1982).

235. Robinson, *supra* note 5, at 747.

236. Delgado, *supra* note 5, at 895.

meet the more-likely-than-not test would actually do them little good, and what relief it did provide would come at the expense of other plaintiffs who would recover more under the traditional test as incorporated in the proposed standard. If in the majority of cases, the attributable risk is above fifty percent, adoption of the theory might do more injustice than justice, even if small impacts were detectable. These potential problems may prove to be more imagined than real, but until they have been carefully researched and considered, the law should not rush to embrace the theory of proportional recovery.[237]

Conclusion

In both toxic tort and cancer cases, courts have generally done a poor job in determining whether evidence of causation is sufficient to meet the plaintiff's burden of proof. Failure to formulate and apply substantive standards has led to irrational and inconsistent results, a problem that need not continue. Accepted legal principles governing the burden of proof, combined with the principles of epidemiology provide an excellent basis for a standard that rationally conforms to general tort law principles and for which there is ample precedent.

Requiring that a plaintiff's evidence satisfy the postulates of epidemiology would work to the disadvantage of the first victims of a

---

237. For single-factor cases, the traditional preponderance of the evidence rule may, as a practical matter, be the best possible standard. Multiple factor cases, however, present additional issues. Consider a plaintiff who has suffered high level exposures to a number of substances, each produced by a different defendant. If the plaintiff has a disease linked to all the substances, proving by a preponderance of the evidence that any single defendant caused it may be difficult. Depending on the levels of exposure and the relative risks for the substances considered separately, each defendant might escape liability only because of the other defendants.

The alternative liability theory, developed in Summers v. Tice, 33 Cal. 2d 80, 84-86, 199 P.2d 1, 2-5 (1948), would not apply to such a situation unless extended. *See* Molloy & Thomas, *Causation Problems in Design Defect Litigation*, Legal Notes & Viewpoints, Feb. 1983, at 35, 44-45 (the doctrine has only rarely been applied in product liability cases). In many toxic tort cases, strict liability rather than negligence is involved, and it is uncertain if all potential defendants have been included. With few exceptions, the case law indicates that this makes alternative liability inapplicable. Other theories formulated to reach multiple defendants would also not apply. These theories are: (1) "concert of action," requiring a common plan or design, which will not often occur in multiple defendant toxic tort cases involving multiple substances; (2) "enterprise liability," requiring industry-wide standards which, unless a single product such as blasting caps is implicated, is unlikely to be useful, and which would therefore be inapplicable in multiple-substance cases; and (3) "market share liability," which is the theory adopted in Sindell v. Abbott Labs., 26 Cal. 3d 588, 611-13, 607 P.2d 924, 936-38, 163 Cal. Rptr. 132, 144-45 (1980). It is difficult to see how a single relevant market could be defined for multiple substances, and thus how the theory could apply to cases involving more than one substance. Thus, some form of causal apportionment among the defendants might well be the best way to allocate liability in a case involving multiple high-level exposures, though further research is required.

substance that eventually proves to have harmful effects. The magnitude of this problem is not, however, as great as many have assumed, and the legal and institutional reforms appropriate to its solution do not involve reducing the plaintiff's burden of proof. Using an epidemiologic standard could cause a sharp shift from under- to over-compensation, but adoption of a proportional recovery standard to ameliorate this problem could create even more serious problems. Further theoretical development of this concept is necessary before it can be recommended. In any event, consistent and rational resolution of toxic tort claims requires that the law incorporate the principles of epidemiology and that legal reforms conform to epidemiologic reality.

Exhibit L

EPIDEMIOLOGY

Concepts and Methods

William A. Oleckno
Northern Illinois University

WAVELAND PRESS, INC.
Long Grove, Illinois

CHAPTER SEVEN

Association and Causation in Epidemiology

This chapter discusses differences among spurious, noncausal, and causal associations, the various types of causes, and common guidelines used in assessing causation in epidemiologic studies.

Learning Objectives

- Describe and give examples of spurious, noncausal, and causal associations in epidemiology.
- State the common reasons for spurious and noncausal associations, respectively.
- Distinguish among necessary, sufficient, necessary and sufficient, necessary but not sufficient, not necessary but sufficient, and not necessary and not sufficient causes and give examples of each type.
- Describe and give examples of direct and indirect causal associations.
- Briefly describe the causal pie model.
- Discuss six guidelines based on Hill's postulates for judging potential causal associations, including the advantages and limitations of each criterion, respectively.
- Explain the importance of finding causal associations in epidemiology.
- Define predisposing or enabling factors, statistical association, and threshold.

Introduction

As indicated in chapter 1, one of the primary goals of epidemiology is to discover the *causes** of morbidity and mortality in human populations. This goal has immense practical significance for health professionals because a better

*There are many terms relating to or derived from the root term *cause*. These include causation, causality, causal, causative, cause-effect, etiology, and so forth. These terms are not defined separately in this chapter, but each refers to something similar.

understanding of the causes of morbidity and mortality often leads to more effective prevention, treatment, and control measures and consequently to a reduction in disease incidence, prevalence, or severity.

A *statistical association* between a given exposure and outcome is the starting point for consideration of a causal relationship in epidemiology. A **statistical association** implies that the exposure is related to a change in the *probability* of the outcome. It does not automatically mean that the exposure *causes* the outcome.[1] Hence, a frequently cited maxim in introductory statistics courses is: "Association does not necessarily imply causation." In short, statistical associations should not be accepted at face value. They should be examined for alternate explanations before any conclusions are drawn. Even a statistically significant association (chapter 6) does not guarantee that a true association exists, much less that the association is causal. A **causal association** between an exposure and outcome means that a change in the frequency of the exposure in a population *will result* in a change in the frequency of the outcome, even though not every individual with the exposure will change. A statistical association only implies that those with the exposure are more or less likely to develop the outcome.

To summarize briefly, a valid statistical association means it is *more or less likely* that the outcome will occur in the presence of the exposure, while a valid causal association means that changes in the frequency of exposure will result in changes in the frequency of the outcome. It should be noted that a causal association may be positive (the exposure increases the outcome) or negative (the exposure decreases the outcome). In the former case, the exposure is *hazardous*; in the latter case it is *protective*. The remainder of this chapter focuses on examining statistical associations to determine whether or not they are likely to represent causal associations. Many factors must be considered, and any conclusions must be based on an overall assessment of the evidence.

TYPES OF ASSOCIATION

Statistical associations found in epidemiologic studies (e.g., OR = 3.4) can be categorized into three types. These categories are mutually exclusive.

- Spurious associations
- Noncausal associations
- Causal associations

Spurious Associations

Spurious associations are literally *false* associations. Though they may be found in a particular study population, they are probably due to other explanations. Spurious associations usually result from *random error* (chance) or *bias*, which are discussed more fully in chapter 8. For example, as mentioned in chapter 6, an association is generally considered statistically significant if $p \leq 0.05$. This implies that, assuming there is no association, chance is an

unlikely explanation for the finding given the sample size and strength of the association. Nonetheless, we would still predict that as many as five times out of 100 the association could be due to chance alone. Thus, even statistically significant associations that result from well-executed epidemiologic studies can sometimes be spurious. Inderjit S. Thind, for instance, conducted an ecological study of the association between dietary intake and cancer using a sample of 60 countries. He found a number of significant statistical associations, including some that were biologically implausible and which he thought to be spurious. In his discussion of the findings, he reiterated a common concern in broad-based studies where large numbers of statistical tests of significance are performed. Specifically, he cautioned the readers by stating, "The . . . large numbers of correlations . . . with [some] significant associations occurring purely by chance, suggest extreme care in assessing the role of specific dietary items as risk factors and using the results as the basis for public policy."[2(p162)]

Spurious associations may also arise from sources of bias. *Bias*, which is discussed in chapter 8, is a type of systematic (nonrandom) error in the design, conduct, or analysis of epidemiologic studies, such as the use of flawed measurement techniques, differential recall among study and comparison groups, or selection of study and comparison groups that are dissimilar. Bias can be quite insidious. Consider a hypothetical case-control study of the relationship between exposure to low-frequency electromagnetic fields, such as those generated by electric power lines, electric blankets, and electric alarm clocks, and the incidence of childhood leukemia. The cases consist of patients from area hospitals newly diagnosed with childhood leukemia, and the controls are those without leukemia of similar age, sex, and racial/ethnic background who have been randomly selected from the communities served by the hospitals. The parents of cases and controls are then queried about their children's exposure to low-frequency electromagnetic fields. The parents of the cases may be more likely to recall their children's exposures than those of the controls since they are probably more motivated to remember past exposures that might help explain their children's leukemia than are the parents of the controls. If this is true, the study could result in a *spurious* association between exposure to low-frequency electromagnetic fields and the incidence of childhood leukemia.

Noncausal Associations

Noncausal associations are real associations, but they are *not* causal associations. That is, a change in the frequency of the exposure in a population does not necessarily result in a change in the frequency of the outcome. Noncausal associations often result from *confounding*, which is discussed in chapter 8. The association exists because the exposure is associated with another factor that in turn is associated with the outcome. A whimsical example is provided by Max Michael III, W. Thomas Boyce, and Allen J. Wilcox.[3] Dr. Al Betzerov conducted a prospective cohort study to test his hypothesis that gambling

causes cancer. He chose two neighboring states, one where gambling was legal and the other where it was not. He then followed randomly selected samples of subjects from each state matched by age, sex, urban/rural differences, and family income for 10 years. At the conclusion of the study, he noted a statistically significant positive association between gambling and cancer. Specifically, the residents of Nevada had a higher rate of cancer than those from Utah. The association, although real, was *not* one of cause-effect. Unfortunately for Dr. Betzerov, one of the states he chose was Utah. Utah is a state composed of a large number of Mormons, who have very different lifestyles from typical Nevada residents, who are not Mormons. The fact that the Mormon Church requires its adherents to abstain from tobacco and alcohol explains this association. The apparent causal association between gambling and cancer was due to confounding by alcohol and tobacco use, which are higher in Nevada than in Utah. In other words, alcohol and tobacco use are associated with gambling and are directly linked to cancer. Therefore, although gambling itself does not cause cancer, its association with causes of cancer produces a noncausal association with cancer. This type of association has also been referred to by some as a "spurious association" in that it can lead to an erroneous conclusion about cause and effect.

Risk markers, which were referred to in chapter 1, represent noncausal associations. Although these associations result from confounding with actual risk factors, they are still real associations that have practical significance in screening for disease.[4] For example, calcification in the coronary arteries is a risk marker for coronary heart disease. It does not cause the disease, but it is associated with an increased risk of its occurrence. Its role in coronary heart disease is therefore properly classified as noncausal. Nevertheless, screening for coronary calcium has become an increasingly popular, though controversial, method of detecting possible presymptomatic heart disease (see chapter 13).

Noncausal associations can also result when the defined exposure is a consequence of the outcome instead of the other way around. Hypertension, for example, may result from kidney disease. Thus, one may find a statistical association between hypertension and kidney disease, but in this example, hypertension could not be considered a cause of kidney disease because the exposure does not *precede* the outcome and therefore cannot alter its frequency. In this example, kidney disease is a cause of hypertension. This type of hypertension is generally referred to as secondary hypertension to differentiate it from primary hypertension, which can cause kidney disease.

Causal Associations

Causal associations are those in which changes in the frequency of the exposure in a population produce a change in the frequency of the outcome. In epidemiology, we cannot prove causal associations because it is impossible to account for all the other factors that might play some role in an association, especially in observational studies where there may be many unrecognized,

and therefore uncontrolled, variables. Well-designed experimental epidemiologic studies can come much closer to establishing causation than observational studies, but even in these studies there may be other influential factors of which the investigator is unaware. Since no two humans beings are exactly alike in their makeup or reactions to external stimuli, one cannot always be assured that even randomized groups of people are perfectly comparable. Even laboratory experiments with mice rely on well-defined strains to minimize intraspecies differences that can invalidate the results of an experiment.

A given association may not be conclusively spurious, noncausal, or causal. This is because random error can never be completely eliminated as a possible reason for an association in an epidemiologic study, although it can be greatly minimized. Similarly, it would be extremely difficult to discount any possibility of bias in a study. The same can be said for possible confounding. Thus, the job of the epidemiologist is to determine which type of association is more likely, and this is not always an easy task.

Since our main concern is identifying causal relationships when they exist, we need some guidance in determining whether an association is likely or not to be a causal one. In practice, the determination of a causal association is based on a careful review and judgment of all relevant information available, and never on the basis of one or two studies alone, especially observational studies. It is somewhat like trying a criminal case where there are no eyewitnesses to the crime. The prosecutor has to rely on circumstantial evidence to convince a jury beyond a reasonable doubt that the defendant is guilty. It was based on a thorough review of major epidemiologic and nonepidemiologic studies that in 1964 the Surgeon General of the U.S. Public Health Service first concluded that cigarette smoking is a cause of lung cancer.[5] Before discussing some of the guidelines used to assess potential causal associations, it should be worthwhile to first examine the concept of causation in more detail. This is the subject of the following section.

TYPES OF CAUSES

With communicable diseases the concept of causation appears to be relatively straightforward. However, as discussed in chapter 3, this apparent simplicity can be deceiving. Not everyone exposed to *Mycobacterium tuberculosis* (the bacterium implicated in tuberculosis), for example, develops tuberculosis. A number of host and environmental factors must also be considered. Similarly, not everyone exposed to cold germs gets a cold. In fact, the more we learn about causation, the more complex it seems. With many noncommunicable diseases, especially chronic conditions like arthritis, mental illness, Alzheimer's disease, multiple sclerosis, cardiovascular disease, diabetes, and so forth, the causal pathways can be extremely complex. Multifactorial etiology (chapter 2) is the rule rather than the exception for most contemporary health-related problems.

Case 3:16-md-02738-MAS-RLS Document 9732-4 Filed 05/07/19 Page 145 of 200 PageID: 34000

Necessary and Sufficient Causes

To get a better understanding of causation as it is commonly used in epidemiology it is helpful to look at different types of causes.* A **necessary cause** is an exposure that is *required* for a particular outcome to occur. Therefore, it is always associated with the outcome. If the exposure is absent, the outcome cannot occur. A **sufficient cause** is an exposure that by itself will produce a particular outcome, but it may not be the only cause of the outcome. Consequently, the outcome may occur without the exposure if the outcome is also caused by other exposures. These two classifications of causes give rise to four possible combinations,[6] which are shown below in the following 2 × 2 table.

	Necessary	
Sufficient	Yes	No
Yes	A	C
No	B	D

Combination A represents a **necessary and sufficient cause**. This is a cause that is required to produce a particular outcome *and* which is able to cause the outcome by itself. This can be represented by:

$$\text{Exposure X} \rightarrow \text{Outcome Y}$$

where Exposure X is the specified cause, and Outcome Y is the specified outcome.

Necessary and sufficient causes are not very common in the real world. One example of a condition that results from a necessary and sufficient cause is lead poisoning. Exposure to lead is *necessary* to produce lead poisoning, and it is also *sufficient*. The rabies virus might also be considered a necessary and sufficient cause of human rabies. It is *not* essential that a necessary and sufficient cause always produces the outcome. Observations have shown, for example, that not everyone presumably infected with the rabies virus contracts the disease even if they have not been immunized.[7] Nevertheless, anyone who contracts rabies must have the virus (i.e., it is necessary), and no other known cause must be present for the disease to occur (i.e., it is sufficient). It is important to emphasize, however, that as knowledge of disease causation expands, classifications may need to be revised. We may learn in the future, for example,

*The types of causes discussed here and subsequently are assumed to be hazardous rather than protective so as to simplify the discussion.

that some causes thought to be necessary and sufficient would be better classified in another way. At one time many believed that cancer was caused by a single factor, still undiscovered. Today we recognize its multifactorial etiology.

Combination B in the above table represents a **necessary but not sufficient cause.** This is a cause that is required to produce a specified outcome *but* is *not* able to cause the outcome by itself. Other causes are necessary for the outcome to occur. This can be represented by:

Exposure X + Other Causes → Outcome Y

Alcoholism is a disease in which alcohol consumption is a necessary but not sufficient cause of the disease. Alcohol consumption is definitely necessary for alcoholism to develop, but other factors, including genetic, social, behavioral, and environmental factors, also appear to be necessary for the disease to manifest itself.

Combination C represents a **not necessary but sufficient cause.** This is a cause that is *not* required to produce a specified outcome *but* when present is able to cause the outcome by itself. This means that there are other causes of the outcome. A not necessary but sufficient cause may be represented by:

Exposure X → Outcome Y and Exposure Z → Outcome Y

where Exposure Z is some other independent cause of Outcome Y. Ionizing radiation at high doses will cause sterility in men. Heavy exposure to certain pesticides will do the same. In this example, Exposure X is ionizing radiation, Exposure Z is a specific pesticide, and Outcome Y is sterility in men. Thus, sterility in men has more than one cause. Both ionizing radiation and certain pesticides are capable of causing sterility in men (at high doses).

Combination D denotes a **not necessary and not sufficient cause.** This is a cause that is *not* required to produce the specified outcome *and* when present is *not* able to cause the outcome by itself. Hence, there are other causes of the specified outcome. A not necessary and not sufficient cause is known as a **contributory cause.** It can be represented by:

Exposure X + Other Causes → Outcome Y and Exposure Z → Outcome Y

where Exposure Z is another independent cause of Outcome Y. Not necessary and not sufficient causes are very common causes of chronic diseases. For example, a sedentary lifestyle is not necessary and not sufficient to cause coronary heart disease (CHD). It is not required for CHD development, nor is it considered sufficient to cause CHD by itself. It is, however, a contributory cause of CHD, and when present with certain other contributory causes, such as high blood cholesterol, family history of heart disease, hypertension, cigarette smoking, and so forth, can lead to the development of CHD. That is, the frequency of CHD will be higher in groups with these factors than in groups without them.

A logical extension of this paradigm is one conceptualized by Kenneth J. Rothman and referred to as the **causal pie model.**[8] One can imagine one or

more intact pies neatly divided into several pieces symbolizing what Rothman calls **component causes**. Each pie represents a *sufficient cause* of a particular disease, and each component cause has an essential part in causing that disease. There may be several sufficient causes (pies) made up of various combinations of some of the same and different component causes for any given disease. Whatever the combination, the component causes work together to cause the disease.[8] The causal pie model may remind one of the information asked for on a death certificate regarding the causes of death (see exhibit 2-1 in chapter 2). In a sense, the immediate, antecedent, and underlying causes of death, as well as other significant conditions, seem to parallel the component causes for a particular death.

As intimated earlier, in epidemiology causation is determined by what occurs in populations or groups of people as opposed to what occurs in any particular individual. We know, for example, based on the Framingham Heart Study that people who live certain lifestyles die more frequently from coronary heart disease than those with healthier lifestyles. From the group data, we can make predictions about individuals based on their lifestyle habits, but we cannot expect that the predictions will always be correct. Everyone seems to know someone, for example, who smoked four packs of cigarettes a day, had high blood pressure, and drank like a fish, but lived until 105. Undoubtedly, this person met an "untimely" death when his bungee cord broke after jumping off a bridge. The exception, however, does not make the rule.

Direct and Indirect Causes

Causal associations can also be classified as direct or indirect. A **direct causal association** (or **direct cause**) can be thought of as representing a causal pathway in which there are *no* intermediate variables, while an **indirect causal association** (or **indirect cause**) involves one or more intervening factors.[9] For example, in a direct causal association, X causes Y, where X is the causative exposure, and Y is the outcome. In an indirect causal association, I causes X, which in turn causes Y. While I is a direct cause of X, it is an *indirect* cause of Y. Since I causes X, and X causes Y, it follows that I causes Y based on the definition of a causal association. A change in the frequency of I in a population will result in a change in the frequency of X, which in turn will result in a change in the frequency of Y. Thus, I can be considered an indirect cause of Y.

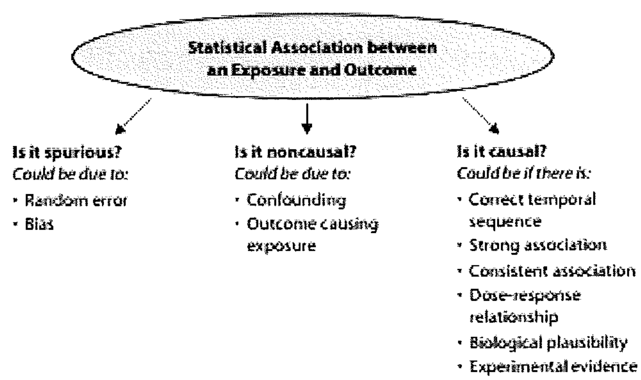
Indirect causes can include a variety of **predisposing or enabling factors** that precede the direct cause. For example, excessive heat applied to the skin is the direct cause of burns, but the exposure to the heat may be influenced by a dangerous working environment or failure to follow certain safety precautions, which might be considered indirect causes of burns. Also, the human immunodeficiency virus (HIV) is said to be the direct cause of AIDS, but factors that facilitate contracting HIV include sharing syringes and promiscuous sexual behaviors. In practice, controlling the predisposing or enabling factors should result in a decrease in frequency of the outcome. Therefore, *predisposing or enabling factors* are often referred to as risk factors.

Whatever classification scheme is used, most contemporary health-related problems appear to have multiple causes. This multifactorial etiology, which has been referred to often in this text, presents a challenge to epidemiologists who are concerned with unraveling the determinants of morbidity and premature mortality and to those whose efforts are directed toward their prevention and control. As our knowledge of the natural history of health problems expands, the models of causation and the methods of intervention will continue to undergo change. An interesting article dealing with different conceptions of casusation from an epidemiologic and philosophical perspective is one published in the *Journal of Epidemiology and Community Health* by M. Parascandola and D. L. Weed.[10] While their recommendations may be at odds with many epidemiologists, the discussion itself is can be enlightening, especially for those new to this topic.

GUIDELINES FOR ASSESSING CAUSATION

As shown in figure 7-1, determining whether a statistical association is causal, involves a number of considerations. One must ask if the observed association is likely to be spurious. Random error or bias could explain an association found in a study population. On the other hand, the association could be a noncausal association. Noncausal associations may be due to confounding by an extraneous factor or because the outcome is responsible for the exposure instead of vice versa. Of course, another option is that the association is causal. Okay, you may say, we know the options, but how can we tell if the association is likely to be a causal one? The first step is to examine whether the alternate explanations are plausible. Specifically, is the associa-

Figure 7-1 Deciding Whether an Association Is Likely to Be Causal



tion likely due to random error, bias, confounding, or a reserved causal sequence? This may take some critical thinking, further analysis, or consultation. If these seem to be unlikely explanations, it can be helpful to review some generally accepted guidelines for establishing causation such as those described by Sir Austin Bradford Hill.

In 1965, Sir Austin Bradford Hill, Professor Emeritus of Medical Statistics with the University of London, delivered a landmark address where he outlined nine criteria that could be used to determine if statistical associations were likely to represent causal associations.[11] His reasoning built on the earlier work of others, such as John Stuart Mill, who in 1856 had defined several canons from which causal relationships could be deduced.[6] Over the years many authors have articulated or modified Hill's basic criteria, which have become known as **Hill's postulates**. Using these as a focal point, the following six guidelines should be helpful in deciding whether or not statistical associations are likely to represent causal associations (figure 7-1). In the end, the process of determining causation is largely subjective except for the first guideline, which is actually a requirement.

- **Correct temporal sequence.** In order for an exposure to be considered a cause of an outcome, it must *precede* the outcome. Of all the guidelines used to judge whether an association is causal or not, this is the only one that is considered *absolutely essential*. Exposures that occur concurrently with an outcome or subsequent to an outcome cannot be considered causal because they do not alter the frequency of the outcome. Determining if an exposure precedes an outcome can be problematic in cross-sectional studies where exposure and outcome are assessed concurrently. For example, in a cross-sectional study designed to determine if there is a relationship between the prevalence of excess body weight and osteoarthritis, it may not be clear which factor came first. Thus, the correct temporal sequence cannot be established reliably. This can also be a problem in case-control studies where the prevalence of the outcome is assessed instead of its incidence.
- **Strength of the association.** In general, the stronger an association between a given exposure and outcome (see table 6-3), the more likely the association is causal. When the risk ratio is very high, for example, it is more difficult to explain away the association due to unrecognized or subtle sources of bias or confounding. Compared to nonsmokers, those who smoke and are exposed to high levels of asbestos in their jobs have a fifty- to ninety-fold increased risk of lung cancer. It seems improbable that these factors are not causative. Even if some bias or confounding exists, it is unlikely that it would account for the entire relationship. This is not to say that small associations cannot also be causal in nature. This is one reason why several guidelines are needed to assess causality.
- **Consistency of the association.** When other investigators studying different populations at different times in different places using different methodologies obtain similar findings with regard to a specific association, it

increases the probability that the association is causal. In concluding that cigarette smoking is a cause of lung cancer, the Advisory Committee to the Surgeon General of the United States cited diverse epidemiologic and other studies showing a strong relationship between smoking and lung cancer.[5] One way of determining if an apparent association is likely to be due to random error is to replicate the study. If the findings are consistent, it strengthens the case for a causal association, assuming there are no significant sources of bias or confounding in the studies.

- **Dose-response relationship.** In general, if increased levels of exposure lead to greater frequencies of the outcome, then this is suggestive of a causal relationship. Heavy smokers, for example, have been shown to be at a higher risk of lung cancer than light smokers. In fact, a linear dose-response relationship between smoking and lung cancer can be demonstrated based on the number of cigarettes smoked per day. The absence of a dose-response relationship does not necessarily mean that an association is noncausal, however. A threshold may exist. A **threshold** is a level of exposure (dose) that must be reached before effects become apparent. Below the threshold, there are no observed effects. Copper, which may be found in small quantities in drinking water and certain foods, demonstrates a threshold; that is, copper has no adverse effects until it reaches a certain level in the body. In fact, in very small quantities it is an essential mineral needed for proper growth and development. On the other hand, a dose-response relationship could be due to a strong confounding factor that closely follows an exposure.[12] Once again, several guidelines should be considered in assessing causation.
- **Biological plausibility.** The basic question here is, does the association make biological sense? Is the association credible based on our understanding of the natural history of the disease or possible pathogenic mechanisms? When Thind found significant associations for protein, fat, and caloric intake and certain forms of leukemia, he could offer no biological evidence to support the associations, thereby casting doubt on their authenticity.[2] Failure to make biological sense, however, does not necessarily negate the possibility of a causal association. In some cases, our understanding of the biological mechanisms may be incomplete, and what does not make sense today may make sense sometime in the future. From a contemporary vantage point, it seems difficult to understand why the theory of contagion was considered controversial as an explanation for the spread of epidemics during the Middle Ages.
- **Experimental evidence.** Having experimental evidence to support an association between a given exposure and outcome strengthens the case for a causal association. Well-designed randomized controlled trials, for example, can provide strong corroboration of a suspected causal association. This is because this study design, properly implemented, can virtually eliminate selection bias and confounding as alternate explanations for a causal

association (see chapters 8 and 12). Of course, the degree of control possible in epidemiologic experiments is not to the same level as that in animal studies. Nevertheless, they can be powerful tools for establishing causation. Evidence from nonepidemiologic experiments can also be used in assessing cause-effect relationships. Because of the limited circumstances in which experimental studies can be conducted with humans, some associations will not be testable in this manner. We would not perform a randomized controlled trial on the effects of microwave radiation on cataract development, for example, because such a study would be unethical even if some were willing to volunteer for the investigation.

Table 7-1 ranks the most common types of epidemiologic studies in descending order of the degree to which identical findings of a statistical association are likely to demonstrate a causal association. The ranking is based on the relative probability of encountering unrecognized bias, confounding, or other errors within the specific study designs. It also assumes that the studies have been planned appropriately and conducted to minimize errors. A poorly designed experimental study can provide less convincing evidence of causality than a well-designed observational study. It should be kept in mind, however, that causality is never determined based on the findings of one study alone. Causation is a judgment based on relevant, cumulative information. Meta-analyses (chapter 12) have provided some hope of reaching more definitive conclusions in epidemiologic studies. Whether they will fulfill this hope depends on the care in which they are designed, implemented, and interpreted.

Table 7-1 Ranking of Common Epidemiologic Studies in Terms of the Relative Probability that the Findings Represent Causal Associations

1. Randomized Controlled Trial	5. Case-Control Study
2. Group Randomized Trial	6. Cross-Sectional Study
3. Prospective Cohort Study	7. Ecological Study
4. Retrospective Cohort Study	8. Descriptive Study

Summary

- Statistical associations found between given exposures and outcomes can be of three types—spurious, noncausal, or causal. Spurious associations are false associations that are usually due to random error or bias. Noncausal associations usually result from confounding, although they can also occur when the exposure is the result of the outcome instead of the other way around. Risk markers represent noncausal associations that have practical value in screening for disease. Causal associations are ones in which a change in the frequency of the exposure results in a change in the frequency of the outcome in a population.

Case 3:16-md-02738-MAS-RLS Document 9732-4 Filed 05/07/19 Page 152 of 200 PageID: 34007

- Causes can be classified as to whether or not they are necessary and/or sufficient and whether they are direct or indirect. A necessary cause is one that is required to produce an outcome, while a sufficient cause is one that can produce the outcome by itself (i.e., in the absence of other known causes). The most common types of causes are those that are not necessary and not sufficient. These are known as contributory causes and are the causes that account for most contemporary health-related problems. The causal pie model expands upon the not necessary and not sufficient causes by considering a constellation of component causes that are sufficient to cause disease. Direct causes do not involve any intermediate factors in the causal pathway. Indirect causes include a variety of predisposing or enabling factors that precede the direct cause of an outcome. Controlling indirect causes can reduce the incidence of particular outcomes and is sometimes easier than controlling the direct causes.
- Because it is not possible to prove causation directly, it is helpful to have reliable guidelines upon which to judge a statistical association in terms of its likelihood of being causal. A final decision regarding causation should be based on all relevant information and not just on the basis of one or two studies, especially observational studies. Six guidelines, derived from Hill's postulates, should help in determining whether an association is likely to be causal. These guidelines are correct temporal sequence, strength of the association, consistency of the association, dose-response relationship, biological plausibility, and experimental evidence. Of these guidelines, only correct temporal sequence is required for an association to be considered causal. The others are highly suggestive of causation, however, especially when all or most of them are met.

New Terms

- biological plausibility
- causal association
- causal pie model
- component causes
- consistency of the association
- contributory cause
- correct temporal sequence
- direct causal association
- direct cause
- dose-response relationship
- experimental evidence
- Hill's postulates
- indirect causal association
- indirect cause
- necessary and sufficient cause
- necessary but not sufficient cause
- necessary cause
- noncausal association
- not necessary and not sufficient cause
- not necessary but sufficient cause
- predisposing or enabling factors
- spurious association
- statistical association
- strength of the association
- sufficient cause
- threshold

Study Questions and Exercises

1. For each of the following statements indicate whether the results are more likely to be due to a spurious association, a noncausal association, or a causal association. Also, explain the reasons for your answers.
   a. A case-control study revealed that there was a moderate to strong association between coffee consumption and deaths from coronary heart disease. Other studies have shown that those who drink coffee are more likely to smoke than those who do not drink coffee.
   b. A prospective cohort study showed that women who exercise regularly were less likely to contract cancer than women who exercised only occasionally or not at all. The exercise group was selected from women attending a fitness center, and the comparison group was selected from women attending a weight-loss clinic.
   c. A large randomized controlled trial showed that folic acid supplementation by prospective mothers significantly reduced the incidence of neural tube defects in their offspring. This finding was confirmed in subsequent studies.
   d. A large exploratory epidemiologic study examined the possible relationship of 25 different lifestyle behaviors to teenage suicide. One of the findings was a positive association between bicycle helmet use and suicide ($p = 0.05$) that had not been previously reported in the literature.
2. On bottles of wine and other alcoholic beverages, it states, "According to the Surgeon General, women should not drink alcoholic beverages during pregnancy because of the risk of birth defects." Discuss the evidence that alcohol consumption causes birth defects using the six guidelines for causation discussed in this chapter. For each guideline, describe the degree to which the evidence supports a conclusion of causation and the reasons for your response. In answering this question it may be necessary to consult a review of epidemiologic literature on alcohol consumption and birth defects.
3. Provide an example other than one used in this chapter of a necessary and sufficient cause, a necessary but not sufficient cause, a not necessary but sufficient cause, and a not necessary and not sufficient cause of disease, respectively. Also indicate why your examples are appropriate.
4. Give two examples, respectively, of direct and indirect causes of disease and justify your choices.

References

1. Vogt, W. P. (1999). *Dictionary of Statistics and Methodology: A Nontechnical Guide for the Social Sciences*, 2nd ed. Thousand Oaks, CA: Sage Publications.
2. Third, I. S. (1986). Diet and Cancer—An International Study. *International Journal of Epidemiology* 15(2): 160–162.

3. Michael, M. III, Boyce, W. T., and Wilcox, A. J. (1984). *Biomedical Bestiary: An Epidemiologic Guide to Flaws and Fallacies in the Medical Literature.* Boston: Little, Brown, and Company.
4. Szklo, M., and Nieto, F. J. (2000). *Epidemiology: Beyond the Basics.* Gaithersburg, MD.: Aspen Publishers, Inc.
5. U.S. Department of Health, Education, and Welfare (1964). *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service.* USPHS Publication No. 1103. Washington, DC: U.S. Government Printing Office.
6. Last, J. M., ed. (2001). *A Dictionary of Epidemiology*, 4th ed. New York: Oxford University Press.
7. Chin, J., ed. (2000). *Control of Communicable Diseases Manual*, 17th ed. Washington, DC: American Public Health Association.
8. Rothman, K. J. (2002). *Epidemiology: An Introduction.* New York: Oxford University Press.
9. Jekel, J. F., Elmore, J. G., and Katz, D. L. (1996). *Epidemiology, Biostatistics, and Preventive Medicine.* Philadelphia, PA: W. B. Saunders Company.
10. Parascandola, M., and Weed, D. L. (2001). Causation in Epidemiology. *Journal of Epidemiology and Community Health* 55: 905-912.
11. Hill, A. B. (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine* 58: 295–300.
12. Brownson, R. C., Remington, P. L., and Davis, J. R. (1998). *Chronic Disease Epidemiology and Control*, 2nd ed. Washington, DC: American Public Health Association.

Exhibit M

Draft Screening Assessment

Talc
($Mg_3H_2(SiO_3)_4$)

Chemical Abstracts Service Registry Number
14807-96-6

Environment and Climate Change Canada
Health Canada

December 2018

Synopsis

Pursuant to section 74 of the *Canadian Environmental Protection Act, 1999* (CEPA), the Minister of the Environment and the Minister of Health have conducted a screening assessment of talc. The Chemical Abstracts Service Registry Number (CAS RN[1]) for talc is 14807-96-6. This substance is among those substances identified as priorities for assessment as it met categorization criteria under subsection 73(1) of CEPA.

Talc is a naturally occurring mineral. According to information reported under section 71 of CEPA and publically available information, in 2011 talc was manufactured in Canada in quantities ranging between 50 to 75 million kg, and in 2016, approximately 100 million kg of talc was imported. In Canada talc is used in adhesives and sealants; automotive, aircraft, and transportation applications; building and construction materials; ceramics; electrical and electronics; textiles; floor coverings; ink, toner, and colourants; lubricants and greases; oil and natural gas extraction applications; paints and coatings; paper and paper products, mixtures, and manufactured items; plastic and rubber materials; toys, playground, and sporting equipment; and in water treatment. The major uses in Canada align with major global uses of talc. Talc is an ingredient in self-care products and is a permitted food additive. In North America, approximately 3 to 4 % of the talc produced and sold is used in cosmetics. High-purity talc is used in cosmetics, while lower-grade talc is used in commercial applications.

The ecological risk of talc was characterized using the Ecological Risk Classification of Inorganic Substances (ERC-I) approach. The ERC-I is a risk-based approach that employs multiple metrics, considering both hazard and exposure in a weight of evidence. Hazard characterization in ERC-I included a survey of past predicted no-effect concentrations (PNECs) and water quality guidelines, or the derivation of new PNEC values when required. Exposure profiling in ERC-I considered two approaches: predictive modelling using a generic near-field exposure model for each substance, and an analysis of measured concentrations collected by federal and provincial water quality monitoring programs. Modelled and measured predicted environment concentrations (PECs) were compared to PNECs, and multiple statistical metrics were computed and compared to decision criteria to classify the potential for causing harm to the environment. The ERC-I identified talc as having a low potential to cause ecological harm.

Considering all available lines of evidence presented in this draft screening assessment, there is a low risk of harm to the environment from talc. It is proposed to conclude that talc does not meet the criteria under paragraphs 64(*a*) or (*b*) of CEPA as it is not entering the environment in a quantity or concentration or under conditions that have or

[1] The Chemical Abstracts Service Registry Number (CAS RN) is the property of the American Chemical Society, and any use or redistribution, except as required in supporting regulatory requirements and/or for reports to the Government of Canada when the information and the reports are required by law or administrative policy, is not permitted without the prior written permission of the American Chemical Society.

may have an immediate or long-term harmful effect on the environment or its biological diversity or that constitute or may constitute a danger to the environment on which life depends.

Talc has been reviewed internationally by other organizations, including the International Agency for Research on Cancer (IARC) and the Danish Environmental Protection Agency. These assessments informed the human health risk assessment.

No critical health effects were identified via the oral or dermal routes of exposure. As such, oral exposure to talc resulting from food intake and self-care products is not of concern. Inhalation exposure from industrial and commercial uses of talc was not identified to be of concern for human health given the limited number of sites producing and processing talc in Canada. Rather, the focus of the assessment is on inhalation and perineal exposure to certain self-care products containing cosmetic- or pharmaceutical-grade talc.

With respect to inhalation exposure, non-cancer lung effects were identified as a critical health effect for risk characterization on the basis of United States National Toxicology Program studies conducted with rats and mice exposed to cosmetic-grade talc. There is potential for inhalation exposure to talc powder during the use of certain self-care products (e.g., cosmetics, natural health products, non-prescription drugs formulated as loose powders). Self-care products formulated as pressed powders (e.g., face makeup) are not of concern. Margins of exposure between air concentrations following the use of dry hair shampoo and critical lung effects observed in animal studies are considered adequate to address uncertainties in the health effects and exposure databases. Margins of exposure between air concentrations following the use of loose powders (e.g., body powder, baby powder, face powder, foot powder) and critical lung effect levels observed in animal studies are considered potentially inadequate to address uncertainties in the health effects and exposure databases.

The meta-analyses of the available human studies in the peer-reviewed literature indicate a consistent and statistically significant positive association between perineal exposure to talc and ovarian cancer. Further, available data are indicative of a causal effect. Given that there is potential for perineal exposure to talc from the use of various self-care products (e.g., body powder, baby powder, diaper and rash creams, genital antiperspirants and deodorants, body wipes, bath bombs), a potential concern for human health has been identified.

Based on the available information, it is proposed that there is potential for harm to human health in Canada at current levels of exposure. Therefore, on the basis of the information presented in this draft screening assessment, it is proposed to conclude that talc meets the criteria under paragraph 64(*c*) of CEPA as it is entering or may enter the environment in a quantity or concentration or under conditions that constitute or may constitute a danger in Canada to human life or health.

It is therefore proposed to conclude that talc meets one of the criteria set out in section 64 of CEPA.

Talc is proposed to meet the persistence criteria but not the bioaccumulation criteria as set out in the *Persistence and Bioaccumulation Regulations* of CEPA.

Table of Contents

List of Tables

1. Introduction

Pursuant to section 74 of the *Canadian Environmental Protection Act, 1999* (CEPA) (Canada 1999), the Minister of the Environment and the Minister of Health have conducted a screening assessment of talc to determine whether this substance presents or may present a risk to the environment or to human health. This substance was identified as a priority for assessment as it met categorization criteria under subsection 73(1) of CEPA (ECCC, HC [modified 2017]).

The ecological risk of talc was characterized using the Ecological Risk Classification of Inorganic Substances (ERC-I) approach (ECCC 2018). The ERC-I is a risk-based approach that employs multiple metrics, considering both hazard and exposure in a weight of evidence. Hazard characterization in ERC-I included a survey of past predicted no-effect concentrations (PNECs) and water quality guidelines, or the derivation of a new PNEC value when required. Exposure profiling in ERC-I considered two approaches: predictive modelling using a generic near-field exposure model for each substance, and an analysis of measured concentrations collected by federal and provincial water quality monitoring programs. Modelled and measured predicted environmental concentrations (PECs) were compared to PNECs, and multiple statistical metrics were computed and compared to decision criteria to classify the potential for causing harm to the environment.

With respect to human health, this draft screening assessment includes the consideration of information on chemical properties, environmental fate, hazards, uses, and exposures, including additional information submitted by stakeholders. Relevant data were identified up to August 2018. Empirical data from key studies, as well as results from models, were used to reach proposed conclusions. Talc has been reviewed internationally through the International Agency for Research on Cancer (IARC) Monographs Programme, United States Environmental Protection Agency (U.S. EPA), the Joint Food and Agriculture Organization of the United Nations (FAO) and the World Health Organization (WHO) Expert Committee on Food Additives (JECFA) and the Danish Environmental Protection Agency (Danish EPA). Talc was also assessed by the Permanent Senate Commission for the Investigation of Health Hazards of Chemical Compounds in the Work Area (MAK-Commission) in Germany and the Cosmetic Ingredient Review (CIR) Expert Panel. These evaluations and reviews were used to inform the health effects characterization in this screening assessment. This assessment focuses on health effects associated with cosmetic-grade talc and not on potential impurities, such as asbestos. Engineered nanomaterials composed of or containing talc are not explicitly considered in this assessment.

This draft screening assessment was prepared by staff in the CEPA Risk Assessment Program at Health Canada and Environment and Climate Change Canada and the Consumer Product Safety Directorate at Health Canada and incorporates input from other programs within these departments. The ecological portion of the assessment is based on the ERC-I document (published May 11, 2018), which was subject to an external peer review and a 60-day public comment period. The human health portion of

this assessment has undergone external peer review and/or consultation. Comments on the technical portions relevant to human health were received from Ms. Lopez, Ms. Super, and Ms. Jeney of Tetra Tech. Although external comments were taken into consideration, the final content and outcome of the screening assessment remain the responsibility of Health Canada and Environment and Climate Change Canada.

This draft screening assessment focuses on information critical to determining whether substances meet the criteria as set out in section 64 of CEPA by examining scientific information and incorporating a weight of evidence approach and precaution.[2] This draft screening assessment presents the critical information and considerations on which the proposed conclusion is based.

2. Identity of substance

Talc (CAS RN[3] 14807-96-6) is one of the softest naturally occurring minerals, made up of magnesium, silicon, and oxygen (ChemIDplus 1993-). The term talc refers to both the pure mineral and a wide variety of soft, talc-containing rocks that are mined and used for a variety of applications (Kogel et al. 2006). Relatively pure talc ore is also referred to as steatite, and soapstone refers to impure, massive talc rock (Fiume et al. 2015).

The mineral talc is composed of triple-sheet crystalline units, consisting of two silicate sheets composed of SiO_4 tetrahedra joined by edge-link $MgO_4(OH)_2$ (Zazenski et al. 1995). These layers, held together loosely via van der Waals forces, slide over one another easily, giving talc its slippery feel and accounting for its softness (Fiume et al. 2015). The size of an individual talc platelet (i.e., a few thousand elementary sheets) can vary from approximately 1 µm to over 100 µm, depending on the conditions of formation of the deposit (Eurotalc 2017). The individual platelet size determines the lamellarity of a sample of talc. Highly lamellar talc will have large individual platelets, whereas microcrystalline talc will have small platelets. Other inorganics in place of magnesium and silicon are common in talc; for example, aluminum and iron may substitute for silicon in the tetrahedral sites, or manganese may substitute for magnesium in the octahedral positions (Zazenski et al. 1995).

---

[2] A determination of whether one or more of the criteria of section 64 of CEPA are met is based upon an assessment of potential risks to the environment and/or to human health associated with exposures in the general environment. For humans, this includes, but is not limited to, exposures from ambient and indoor air, drinking water, foodstuffs, and products available to consumers. A conclusion under CEPA is not relevant to, nor does it preclude, an assessment against the hazard criteria specified in the *Hazardous Products Regulations*, which are part of the regulatory framework for the Workplace Hazardous Materials Information System for products intended for workplace use. Similarly, a conclusion on the basis of the criteria contained in section 64 of CEPA does not preclude actions being taken under other sections of CEPA or other acts.

[3] The Chemical Abstracts Service Registry Number (CAS RN) is the property of the American Chemical Society and any use or redistribution, except as required in supporting regulatory requirements and/or for reports to the Government of Canada when the information and the reports are required by law or administrative policy, is not permitted without the prior written permission of the American Chemical Society.

Commercially exploited talc contains 20 to 99 % of the pure mineral (Kogel et al. 2006). Some of the most common minerals that occur with talc are carbonates (e.g., dolomite, calcite, magnesite) and chlorite (i.e., magnesium aluminum silicate) (CIR 2013). Less common minerals include quartz, mica, iron oxides, pyrite, serpentine, and amphibole. Selective mining, ore processing, and beneficiation can remove many of the impurities (Kogel et al. 2006). There is a trend towards upgrading and higher-purity talc; however, many applications require the properties of the minerals associated with talc (Kogel et al. 2006). The purity of the source talc will influence its uses.

There are different grades of talc that refer to the purity (presence of other minerals). Pharmaceutical-grade talc conforms to the United States Pharmacopeia (USP) specifications (or similar specifications); these specifications require the absence of asbestos and set limits on iron, lead, calcium, and aluminum (USP 2011). As per B.01.045 of the *Food and Drug Regulations*, when used as a food additive talc must comply with Food Chemical Codex specifications or the Combined Compendium of Food Additive Specifications, prepared by the Joint FAO/WHO Expert Committee on Food Additives, and must be free from asbestos (FAO 2006).

Cosmetic-grade talc should comply with USP standards that require a limit of 20 ppm lead and an absence of asbestos (Fiume et al. 2015). Historically, some talc source materials were contaminated with asbestos; however, in 1976 the Cosmetic Toiletry Fragrance Association (CTFA) set purity standards for cosmetic-grade talc (Fiume et al. 2015). In Canada, the *Prohibition of Asbestos and Products Containing Asbestos Regulations* to be made under CEPA 1999 will prohibit asbestos above trace levels in consumer products, including cosmetics. Health effect studies on cosmetic-grade talc cited in this assessment were considered to be free of asbestos.

Talc is milled to different particle sizes for specific commercial applications. Most talc for cosmetics and pharmaceuticals are pure 200-mesh roller-milled talc (Kogel et al. 2006). In 200-mesh talc (preferred for body powder and deodorants), the particle size distribution allows 95 to 99 % of the product to pass through a 200-mesh (74 μm) screen (Zazenski et al. 1995; Kogel et al. 2006). The finer 325-mesh talc is also used in cosmetic-, pharmaceutical-, and food-grade formulations, where 95 to 99 % of the product passes through a 325-mesh (44 μm) screen.

3. Physical and chemical properties

A summary of physical and chemical properties of talc is presented in

Table 3-1. Talc is hydrophobic and lipophilic (Kogel et al. 2006).

Table 3-1. Experimental physical and chemical property values (at standard temperature) for talc

Property	Range	Key reference
Physical state	solid, powder	HSDB 2005
Melting point (°C)	1500	Eurotalc 2017
Vapour pressure (mm Hg)	approx. 0, negligible at 20°C	OSHA 1999; NIOSH 2014
Water solubility (mg/L)	insoluble	HSDB 2005
Specific gravity (unitless)	2.58–3.83	HSDB 2005

4. Sources and Uses

Talc is a naturally occurring mineral, and there are deposits of talc in most provinces of Canada (Kogel et al. 2006). Currently, there is one producing mine (open-pit) and concentrator facility in Canada, in Penhorwood Township near Timmins, Ontario, and one micronizing facility in Timmins (Kogel et al. 2006; MAC 2016; NPRI 2018). The talc ore from the mine is approximately 45 % pure, with magnesite, magnetite, chlorite, and serpentine as the major impurities (Kogel et al. 2006). After beneficiation, this mine and micronizing facility produces talc primarily for the paper, plastics, paint, and ceramic sectors (Kogel et al. 2006). In 2017, China was the largest producer of talc, followed by India, Brazil, Mexico, and Korea (USGS 2018). The major uses of talc globally include paper, plastics, paint, ceramics, putties, and cosmetics (USGS 2000; Kogel et al. 2006; EuroTalc 2017; USGS 2018) and are aligned with Canadian uses.

On the basis of information submitted pursuant to a CEPA section 71 survey for the year 2011, talc was reported to be manufactured and imported in Canada at quantities ranging from 50 to 75 million kg (EC 2013).[4] According to the Canadian International Merchandise Trade (CIMT) database, in 2016, 99 549 000 kg of natural steatite and talc, crushed or powdered (Harmonized System, HS code 252620) and 4 656 000 kg of natural steatite and talc, not crushed, not powdered (HS code 252610) were imported into Canada (CIMT 2017).

According to information reported pursuant to a CEPA section 71 survey, results from voluntary stakeholder engagement (ECCC, HC 2017), and a search of websites from talc producers, manufactured or imported talc is used in Canada in: adhesives and sealants; automotive, aircraft, and transportation applications; building and construction materials (e.g., wood and engineered wood); ceramics; electrical and electronics; textiles; floor coverings; ink, toner, and colourants; lubricants and greases; oil and natural gas extraction applications; paints and coatings; paper and paper products,

[4] Values reflect quantities reported in response to the survey conducted under section 71 of CEPA (EC 2013). See survey for specific inclusions and exclusions (schedules 2 and 3).

mixtures, or manufactured items; plastic and rubber materials; toys, playground, and sporting equipment; and in water treatment.

Talc is a formulant in pest control products registered in Canada (Health Canada 2010, Personal communication, email from the Pest Management Regulatory Agency, Health Canada to the Risk Management Bureau, Health Canada, dated March 29, 2017; unreferenced).

Additionally, in Canada talc is on the List of Permitted Food Additives with Other Accepted Uses for limited uses in a small number of foods (Health Canada [modified 2017]). Talc can be used as a coating agent on dried legumes and rice and as a filler and dusting powder for chewing gum as per the List of Permitted Food Additives with Other Accepted Uses, incorporated by reference into its respective Marketing Authorization issued under the *Food and Drugs Act*. It may be present in food packaging materials and in incidental additives[5] used in food processing establishments (email from the Food Directorate, Health Canada, to Existing Substances Risk Assessment Bureau, Health Canada, dated March 31, 2017; unreferenced).

Talc is present in approximately 8500 self-care products.[6] Talc is marketed or approved as a non-medicinal ingredient in approximately 1600 human and veterinary drug products in Canada, including approximately 150 over-the-counter (OTC) or non-prescription products (email from the Therapeutic Products Directorate, Health Canada, to the Existing Substances Risk Assessment Bureau, Health Canada, dated March 20, 2017; unreferenced). Talc is listed in the Natural Health Products Ingredients Database (NHPID [modified 2018]) with a medicinal role and classified as a natural health product (NHP) substance falling under item 7 (a mineral) of Schedule 1 to the *Natural Health Products Regulations* and with a non-medicinal role (NHPID [modified 2018]). Talc is listed in the Licensed Natural Health Products Database (LNHPD) as being present as a medicinal or non-medicinal ingredient, in currently licensed natural health products in Canada (LNHPD [modified 2018]). Talc is present as a medicinal or a non-medicinal ingredient in approximately 2000 active licensed NHPs. Talc is listed as a medicinal ingredient in diaper rash products in concentrations ranging from 45 to 100 % in the Diaper Rash Monograph (Heath Canada 2007); however, there are no diaper rash products listed in the LNHPD containing talc as a medicinal ingredient (LNHPD [modified 2018]). Talc is permitted as a medicinal ingredient in the monograph for Traditional Chinese Medicine Ingredients (Health Canada 2015).

[5] While not defined under the Food and Drugs Act (FDA), incidental additives may be regarded, for administrative purposes, as those substances that are used in food processing plants and that may potentially become adventitious residues in foods (e.g., cleaners, sanitizers).

[6] Self-care products are products available for purchase without a prescription from a doctor, and fall into one of three broad categories: cosmetics, natural health products, and non-prescription drugs.

Based on notifications submitted under the *Cosmetic Regulations* to Health Canada, talc is an ingredient in approximately 6500 cosmetic products in Canada (dated April 5, 2017, emails from the Consumer Product Safety Directorate, Health Canada, to the Existing Substances Risk Assessment Bureau, Health Canada; unreferenced). Talc is considered a restricted ingredient in cosmetics.[7] The Cosmetic Ingredient Hotlist entry for cosmetics containing talc in powder form intended to be used on infants and children indicates that product labels should display text to the effect of “keep out of the reach of children” and “keep powder away from child’s face to avoid inhalation that can cause breathing problems.” High-purity talc (fewer impurities of other minerals) is used in cosmetics, while lower-grade talc is used in the many commercial applications mentioned above. In North America, approximately 3 to 4 % of the talc produced and sold is used in cosmetics (Kogel et al. 2006; USGS 2018).

Condoms and medical gloves are regulated as Class II medical devices in Canada under the *Medical Devices Regulations* and may be sources of exposure if talc is present as a dry lubricant. However, a 1998 study did not find talc in a small survey of condoms tested in Canada (Douglas et al. 1998). Condom standards require dry lubricants to be bioabsorbable, such as starch and calcium carbonate (WHO, UNFPA, FHI 2013). Starch is more commonly used as dry powder lubricant on condoms (Douglas et al. 1998). There was also a shift from the use of talc as a dry lubricant on medical patient examination gloves to cornstarch in the 1980s (Lundberg et al. 1997). In 2016, the U.S. Food and Drug Administration banned powdered patient examination gloves (United States 2016).

5. Potential to cause ecological harm

5.1 Characterization of ecological risk

The ecological risk of talc was characterized using the Ecological Risk Classification of Inorganic Substances (ERC-I). The ERC-I is a risk-based approach that employs multiple metrics that consider both hazard and exposure in a weight of evidence. Hazard characterization in ERC-I included a survey of past domestic and international assessment PNECs and water quality guidelines. When no suitable existing PNEC or water quality guideline was found, hazard endpoint data were collected and, dependent on data availability, either a species sensitivity distribution (SSD) or an assessment factor (AF) approach was taken to derive a new PNEC value. In the case of talc, hazard endpoint data from the Organisation for Economic Co-operation and Development

[7] Talc is described as a restricted ingredient on the List of Prohibited and Restricted Cosmetic Ingredients (more commonly referred to as the Cosmetic Ingredient Hotlist or simply the Hotlist), an administrative tool that Health Canada uses to communicate to manufacturers and others that certain substances may contravene the general prohibition found in section 16 of the *Food and Drugs Act* (FDA), or may contravene one or more provisions of the *Cosmetic Regulations*. Section 16 of the FDA states that "no person shall sell any cosmetic that has in or on it any substance that may cause injury to the health of the user." In addition, the Hotlist includes certain substances that may make it unlikely for a product to be classified as a cosmetic under the FDA (Health Canada [modified 2018]).

Screening Information Dataset (SIDS) for synthetic amorphous silicates (OECD 2004) were identified for read across (ECCC, HC 2017) and an AF approach was used to derive a PNEC value of 40 mg/L.

Exposure profiling in ERC-I considered two approaches: predictive modelling using a generic near-field exposure model, and an analysis of measured concentrations collected by federal and provincial water quality monitoring programs. The generic near-field exposure model used input data, when available, from the National Pollutant Release Inventory (NPRI), the DSL–Inventory Update (DSL-IU), international trade data from the Canada Border Services Agency (CBSA), and third-party market research reports to generate PECs. In the case of talc, input data from the DSL-IU and CBSA were available.

Modelled PECs were compared to PNECs, and statistical metrics considering both the frequency and magnitude of exceedances were computed and compared to decision criteria to classify the potential for ecological risk as presented in ECCC (2018). The results are summarized in Table 5-1. The ERC-I identified talc as being of low ecological concern.

Table 5-1. Ecological risk classification of inorganics results for talc

Monitoring (total/extractable)	Monitoring (dissolved)	Modelling (DSL-IU)	Modelling (NPRI)	Modelling (CBSA)	Overall ERC-I score
NA	NA	Low	NA	Low	Low

Abbreviations: NA, Not Available.

6. Potential to cause harm to human health

6.1 Health effects assessment

Talc was previously reviewed internationally by the IARC, and an IARC monograph is available (IARC 2010). Additionally, talc was reviewed by the United States Environmental Protection Agency (U.S. EPA), the Joint FAO/WHO Expert Committee on Food Additives (JECFA), the Permanent Senate Commission for the Investigation of Health Hazards of Chemical Compounds in the Work Area (MAK-Commission) in Germany, and the Danish Environmental Protection Agency (Danish EPA) (U.S. EPA 1992; JECFA 2006; MAK-Commission 2012; Danish EPA 2016). Talc's safety in cosmetic uses was also assessed by the CIR Expert Panel (CIR 2013; Fiume et al. 2015).

A literature search was conducted from the year prior to the most recent assessment (the 2016 Danish EPA review), i.e., from January 2015 to January 2018. No health effects studies that could impact the non-cancer risk characterization (i.e., result in different critical endpoints or lower points of departure than those stated in existing reviews and assessments) for oral, dermal, or inhalation exposures were identified. For perineal exposures, recently published literature was identified and considered in the assessment.

The health effects of talc are outlined by route of exposure in the following sections.

Toxicokinetics

Talc is poorly absorbed via the oral route of exposure. Following gavage administration of radiolabelled talc to rodents, the majority of the administered dose (AD) remained in the gastrointestinal (GI) tract and was eliminated and recovered in the faeces (≥ 95.8 % of AD) within three to four days of dosing (Wehner et al. 1977a; Phillips et al. 1978). Less than 2 % of the AD was recovered in the urine; however, this was mainly attributed to contamination from faeces during collection, with true absorption and urinary clearance expected to be even lower. At 24 hours post administration, less than 2 % of the AD remained in the carcass of hamsters; no radioactivity was detected in mouse carcasses at this time point. In rats and guinea pigs, only trace amounts of radioactivity remained in the GI tract at 10 days post administration.

As an insoluble solid, talc is not expected to be absorbed when applied to healthy and intact skin. There are no indications of dermal absorption following talc exposure (MAK-Commission 2012).

Inhalable talc particles (<10 µm) are eliminated from the respiratory tract via mucociliary clearance. In female Syrian hamsters that were administered aerosolized neutron-activated cosmetic talc at concentrations of 40 to 75 mg/m^3 (95% pure; MMAD 6.4 to 6.9 µm) over a 2-hour exposure period, 6 to 8 % of the AD was deposited into the alveoli (Wehner et al. 1977b). The biological half-life following a single exposure was estimated to be between 7 and 10 days, with complete alveolar clearance after 4 months. There was no translocation of talc from the respiratory tract to the liver, kidneys, ovaries, or other parts of the body. Lung clearance was noted to be longer in other species. The Danish EPA (2016) noted that talc, including the respirable fraction (< 4 µm), is not absorbed following inhalation, but is retained in the lung tissue. They further stated that lung burdens were proportional to respired concentrations, and clearance became impaired with increasing exposures. Pulmonary retention half-lives for talc particles in the lungs of rats from a chronic inhalation study were estimated to be as long as 300 days (Oberdorster 1995). Other authors (Pickrell 1989; MAK-Commission 2012) noted similar findings indicating that with repeat exposures, alveolar clearance in rats may be impaired at concentrations of only 2 mg talc/m^3 air.

Talc particles have been observed and detected in the ovaries of humans (Heller et al. 1996a, 1996b), and perineal exposure to talc has also been associated with a presence of talc in lymph nodes and ovaries of women diagnosed with ovarian cancer (Heller et al. 1996b; Cramer et al. 2007). Migration of talc particles from the vagina to the ovaries has been identified as a plausible explanation of these findings (Henderson et al., 1986), and retrograde movement of talc particles in humans through the reproductive tract to the ovaries has been suggested (Heller et al. 1996b; Cramer et al. 2007). Inert particles with the same size as talc (5 to 40 µm in diameter) and placed in the vagina can be transported to the upper genital tract (Egli and Newton 1961; De Boer 1972; Venter and Iturralde 1979).

According to a review by the MAK-Commission (2012), there are no indications of metabolism via typical degradation pathways from which toxicologically relevant degradation products may develop.

Health Effects

Oral route of exposure

Talc was considered be of low concern with respect to human health via oral exposure. Repeated-dose testing with talc in animals did not produce any adverse effects via oral exposure with respect to repeated-dose toxicity, carcinogenicity, reproductive/developmental toxicity, or mutagenicity (Gibel et al. 1976; Wagner et al. 1977; NTP 1993; IARC 2010; Danish EPA 2016).

Talc has not been shown to produce adverse effects when ingested orally; as a result, the use of talc in various tablet formulations was not considered hazardous via the ingestion route (Hollinger 1990; U.S. EPA 1992).

In addition, the Commission of the European Communities' report on Dietary Food Additive Intake in the European Union identified talc as having an Acceptable Daily Intake (ADI) of "not-specified." The JECFA has also assessed talc and assigned an ADI as "not specified" due to the lack of toxicity from oral exposure. The substance was considered not to be a hazard to human health at oral intake levels noted in total diet surveys, which represent the majority of the sources of oral exposure for this substance (IARC 1987; EU [modified 2001]). Furthermore, talc is considered as "generally recognized as safe" when used as a food additive in the United States (U.S. FDA GRAS list) without being subject to pre-market approval requirements (U.S. FDA 2015; 2016).

Dermal route of exposure

There are limited data available on repeated-dose studies via dermal exposure to talc (Danish EPA 2016). In the available literature, only one repeated-dose dermal toxicity study was identified (Wadaan 2009). Severe limitations were noted for this study, including a lack of information on the test substance and the dose applied, as well as a lack of detail regarding the test animals. Skin dryness and erosion were noted; however, application sites were shaved, indicating that talc may have been applied to broken skin. As such, the results of this study were not considered appropriate to inform the characterization of health effects via dermal exposure. Additionally, there were no indications of irritation, sensitization, or dermal absorption following exposure to unabraded and/or non-diseased skin (MAK-Commission 2012). A three-day occlusive application of pharmaceutical-grade talc did not show any signs of irritation in 5 human volunteers (Frosch and Kligman 1976, as reported in MAK-Commission 2012).

Case reports, however, do indicate that the application of talc to diseased or broken skin can cause the formation of granulomas, particularly if the talc particles have a large diameter (MAK-Commission 2012; CIR 2013; Fiume et al. 2015). Granulomas have

been observed in the umbilical regions of infants, in the testes, on the vocal cords, in the urinary tract, and during phlebectomies following contact with talc-powdered surgical gloves (Ramlet 1991, Simsek et al. 1992, as reported in MAK-Commission 2012). As a result, the CIR concluded that “talc should not be used on skin where the epidermal barrier is removed or on skin that has greater than first degree burns.”

Although dermal contact with talc is expected from the use of various products available to consumers, talc is a solid powder that is insoluble in water (Table 3-1). As a result, it cannot readily penetrate intact skin, and therefore systemic absorption through the skin is not expected. Consistent with other international regulatory and advisory bodies (Danish EPA, U.S. EPA, MAK-Commission, U.S. FDA, and JECFA), a dermal health effects endpoint has not been identified for talc.

Inhalation route of exposure

Human studies

The Danish EPA (2016) noted that talc is not absorbed via inhalation. Rather, particles are retained in the lung, and lung burdens increase proportionally with exposure concentrations or frequency. The report detailed epidemiological data that noted mortalities in workers due to lung diseases, following exposures to talc. However, it was stated that there was no increase in the lung cancer rate in talc millers in the absence of exposure to carcinogens. A recent meta-analysis by Chang and colleagues (2017) reported a positive association with lung cancer in workers exposed to talc; however, co-exposure to other hazardous materials in the workplace and smoking were not adequately accounted for.

The chronic inhalation of talc leads to lung function disorders and fibrotic changes in humans. Since talc particles are persistent, particles accumulate in human lung tissue. This accumulation may lead to both an impairment of the self-purification function (reduced ability to fight infections) and inflammatory changes and fibrosis. Talc particles may be enclosed in a foreign-body granuloma as the result of an inflammatory reaction. The immobility of the macrophages, which is restricted by the phagocytized talc particles, leads to changes in the function of these cells and subsequently to chronic inflammatory reactions (Gibbs et al. 1992).

In humans, there are reports of pure talc-induced pneumoconiosis or talcosis following inhalation exposure to talc. Talcosis has been reported to occur in miners, millers, rubber workers, and other occupational groups exposed to talc without asbestos or silica (Vallyathan and Craighead 1981; Feigin 1986; Gibbs et al. 1992; Akira et al. 2007). Specifically, a recent longitudinal survey of French and Austrian talc workers found that the prevalence of small radiological opacities and decreases in lung function parameters were related to cumulative exposure. The mean estimated talc dust concentration during the mean duration of follow-up (14.5 years) was 1.46 mg/m^3 (Wild et al. 2008). Case reports indicate that patients present with non-specific complaints, including progressive exertional dyspnea, dry or productive cough, with indications of

lung lesions (Marchiori et al. 2010; Frank and Jorge 2011). Talcosis has been shown to occur in children and adults, with symptoms that developed shortly after acute to short-term exposure or up to 10 years later (Patarino et al. 2010; Shakoor et al. 2011). Inhalation of talc has been known to cause pulmonary effects, even following single acute exposures, as reported in a 10-year-old child who had a history of a single exposure to talc at two years of age (Cruthirds et al. 1977). Another case report detailed a seven-year-old child who developed asthma and reduced lung function after a single exposure event (Gould and Barnardo, 1972). Additionally, a 52-year-old woman who used baby talcum powder regularly at least twice a day (usually after bathing for personal hygiene and habitually applying it to her bed sheets nightly) for 20 years was reported to have dyspnea, along with a persistent dry cough and unintentional rapid weight loss. A radiographic exam noted evidence of interstitial lung disease with fibrosis (Frank and Jorge 2011).

Other relevant case reports include the case of a 55-year-old woman, occupationally exposed to talc as a dusting agent on packed rubber balls from 1958 to 1968, who was reported to develop dyspnea during the first five years after exposure (Tukiainen et al. 1984); and a 62-year-old woman occupationally exposed to talc for five years who was reported to have progressive lung fibrosis for more than 40 years (Gysbrechts et al. 1998).

Animal studies

In a repeated-exposure study conducted by the U.S. National Toxicology Program (NTP), groups of F334/N rats were exposed to aerosolized talc via the inhalation route of exposure. Test animals were exposed for 6 hours per day, 5 days per week, for up to 113 weeks (males) or up to 122 weeks (females) to aerosols of 0, 6, or 18 mg/m^3 talc (49 or 50 males per group, 50 females per group) (NTP 1993). Mean body weights of rats exposed to 18 mg/m^3 talc were slightly lower than those of controls after week 65. No clinical observations were attributed to talc exposure. Absolute and relative lung weights of male and female rats exposed to 18 mg/m^3 talc were significantly greater than those of controls. Inhalation exposure produced a spectrum of inflammatory, reparative, and proliferative processes in the lungs. Granulatomous inflammation, which was evident as early as 6 months (first histopathological examination), occurred in nearly all exposed rats, and the severity increased with exposure duration and concentration. Hyperplasia of the alveolar epithelium and interstitial fibrosis occurred in or near the foci of inflammation in many exposed rats, while squamous metaplasia of the alveolar epithelium and squamous cysts were also occasionally seen. Accumulations of macrophages (histiocytes), most containing talc particles, were found in the peribronchial lymphoid tissue of the lung and in the bronchial and mediastinal lymph nodes. In exposed male and female rats, there was a concentration-related impairment of respiratory function, beginning at 11 months, which increased in severity with increasing exposure duration. The impairment was characterized by reductions in lung volume (total lung capacity, vital capacity, and forced vital capacity), lung compliance, gas exchange efficiency (carbon monoxide diffusing capacity), and non-uniform intrapulmonary gas distribution (NTP 1993).

In female rats at 18 mg/m^3 talc, the incidences of alveolar/bronchiolar adenoma, carcinoma, and adenoma or carcinoma (combined) were significantly greater than those of controls (NTP 1993). The incidences of lung neoplasms in exposed male rats were similar to those in controls. Adrenal medulla pheochromocytomas (benign, malignant, or complex [combined]) occurred with a significant positive trend in male and female rats, and the incidences in the 18 mg/m^3 talc groups were significantly greater than those of controls (NTP 1993).

The NTP (1993) concluded that there was some evidence of carcinogenic activity of talc in male rats on the basis of an increased incidence of benign or malignant pheochromocytomas of the adrenal gland. The NTP also concluded that there was clear evidence of carcinogenic activity of talc in female rats on the basis of increased incidences of alveolar/bronchiolar adenomas and carcinomas of the lung and benign or malignant pheochromocytomas of the adrenal gland.

In a subsequent symposium, experts from the NTP, along with academic, industry, and government experts re-examined the results of the chronic inhalation studies. The general consensus from the expert panel was that the highest dose tested (18 mg/m^3) exceeded the Maximum Tolerated Dose (MTD) and as such, the neoplasms noted were not relevant to human health risk assessment (Carr 1995). A similar conclusion was rendered by Warheit et al. (2016). In addition, the Danish EPA (2016) and the MAK-Commission attributed lung tumours in female rats to the general particle effect of granular biopersistent dusts, which manifests as tumours in rodents only, and not the specific effect of the talc particles. They also attributed the pheochromocytomas to an increase in cell proliferation due to hypoxia, which was considered to be a high-dose effect (MAK-Commission, 2012).

A chronic, repeated-exposure study was conducted in B6C3F1 mice via the inhalation route of exposure (NTP 1993). Test animals were exposed for 6 hours per day, 5 days per week, for up to 104 weeks to aerosols of 0, 6, or 18 mg/m^3 talc (47 to 49 males per group, 48 to 50 females per group). Survival and final mean body weights of male and female mice exposed to talc were similar to those of the controls. There were no clinical findings attributed to talc exposure. Inhalation exposure of mice to talc at both concentrations was associated with chronic active inflammation and the accumulation of macrophages, which contained talc, in the lung. In contrast to rats, hyperplasia of the alveolar epithelium, squamous metaplasia, or interstitial fibrosis were not associated with the inflammatory response in mice, and the incidences of lung neoplasms in exposed and control groups of mice were similar. Accumulations of macrophages (histiocytes) containing talc particles were also present in the bronchial lymph node. The critical-effect level and corresponding health effects endpoint was a lowest observed adverse effect concentration (LOAEC) of 6 mg/m^3 for non-cancer lung effects (NTP 1993).

Doses used in the NTP chronic studies were selected on the basis of the results of a 4-week inhalation study (1993) in which rats and mice were exposed to talc at 0, 2, 6, or 18 mg/m^3, 6 hours a day, 5 days a week. Lung burdens were noted to be increased in a

dose-dependent manner, with overload noted by the study authors at 6 and 18 mg/m^3 in rats but not at any dose in mice. In both species (mice and rats), a minor macrophage infiltration of lung tissue was the only health effect noted in the high-dose animals, while animals in the mid- and low-dose groups were without treatment-related effects.

In a review of the NTP studies, Oberdorster (1995) revisited the lung deposition data and particle accumulation kinetics in the lungs of rats and mice in those studies, demonstrating that impaired clearance and lung overload was reached at 6 mg/m^3 and above, for both sexes, in rats and mice.

A no-observed adverse effect concentration (NOAEC) of 2 mg/m^3 was derived from the 4-week study, on the basis of increased lung burden and impaired clearance at a LOAEC of 6 mg/m^3 following 4-weeks of dosing, which led to non-cancer lung lesions at this concentration when the duration of dosing was extended. Granulatomous inflammation and alveolar epithelial hyperplasia were noted at a 6 month interim sacrifice in the chronic rat inhalation study, with interstitial fibrosis and impaired lung function noted in some animals at 11 months. As noted previously, following a single exposure in rats, the biological half-life for ciliary clearance was between 7 and 10 days, indicating that previous exposure would not have cleared prior to subsequent exposures, leading to a build-up in lung tissue. A re-examination of the NTP lung burden data by Oberdorster (1995) estimated that lung retention half-lives of talc particles were between 250 and 300 days in the rat chronic study. On the basis of this information, it was considered relevant to combine the NTP studies for the derivation of an appropriate point of departure for lung effects associated with repeated inhalation exposures.

The Danish EPA (2016) used the LOAEC of 6 mg/m^3 from the chronic NTP studies (mice and rats) and a NOAEC of 1.5 mg/m^3 for talc-induced non-cancer lung effects in the longitudinal survey of French and Austrian talc workers (Wild et al. 2008) to establish a health-based quality criterion for ambient air (QC_{air}) of 0.004 mg/m^3.[8]

While human occupational studies and case studies are available, these studies do not provide accurate measures of exposure for use in risk characterization. However, human studies do note a similar range of lung effects and disease as animal models. As such, results from the animal studies noted above were selected for the non-cancer risk characterization. On the basis of the NTP studies with rats and mice exposed to cosmetic-grade talc, a NOAEC of 2 mg/m^3 for non-cancer lung effects is considered to be appropriate for the inhalation route of exposure for short- or long-term use (given the long half-life and slow lung clearance of talc from the lungs, even episodic exposures would be expected to increase lung load). The NOAEC of 2 mg/m^3 was adjusted according to U.S. EPA guidance on inhalation risk assessment for a comparison with

[8] The health-based quality criterion in ambient air (QC_{air}) is a reference concentration that refers to the maximum permissible contribution to air from industrial sources.

exposure estimates (U.S. EPA 1994, 2009).[9] The adjusted NOAEC for non-cancer effects is 0.36 mg/m^3.

Perineal exposure to talc

The IARC has classified perineal use of talc-based body powder as “possibly carcinogenic to humans” (Group 2B) on the basis of limited evidence in humans. The analyzed case-control studies found a modest but consistent increase in risk, although bias and confounders could not be ruled out. The IARC Working Group concluded that, taken together, the epidemiological studies provide limited evidence in humans of an association between perineal use of talc-based body powder and an increased risk of ovarian cancer, although a minority of the Working Group considered the evidence inadequate because the exposure-response was inconsistent and the cohort analyzed did not support an association (IARC 2010).

The CIR Expert Panel (2013) determined that there is no causative relationship between cosmetic use of talc in the perineal area and ovarian cancer, and further concluded that talc is safe in the practices of use and concentration described in the CIR safety assessment. Issues noted by the CIR included a lack of consistent statistically significant positive associations across all studies; small risk ratio estimates; a failure to rule out other plausible explanations such as bias, confounders, and exposure misclassifications; and a lack of evidence from studies of occupational exposures and animal bioassays (CIR 2013; Fiume et al. 2015).

Animal studies

Rodents are poor experimental models for perineal studies for a number of reasons. Ovulation in rodents occurs only or mainly during the breeding season, and rodent ovaries are variously enclosed in an ovarian bursa in comparison to human ovaries. Ovarian epithelial tumours are also rare in these animals (Taher et al. 2018). Ovarian tumours do occur in some strains of mice and rats; however, the low incidence and/or the length of time required for the appearance of tumours renders them poorly feasible for experimental studies of ovarian carcinogenesis (Vanderhyden et al. 2003). On account of the limitations detailed above, in addition to the challenges posed by exposing animals via the perineal route, animal data are very limited; one single-dose study and one short-term repeated-dose study were available (Hamilton et al. 1984;

[9] This adjustment was made according to guidance and equations outlined in the U.S. EPA Supplemental Guidance for Inhalation Risk Assessment (US EPA 2009) and the U.S. EPA Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry (U.S. EPA 1994). Adjustment of duration to a continuous exposure scenario is done through the use of Equation 1 from U.S. EPA 2009 where the NOAEL[ADJ] = E × D × W, whereby the NOAEL[ADJ] (mg/m^3) = the no-observed adverse effect level (NOAEL) adjusted for the duration of the experimental regimen; E (mg/m^3) = the NOAEL or analogous exposure level observed in the experimental study; D (h/h) = the number of hours exposed/24 hours; and W (days/days) = the number of days of exposure/7 days. The NOAEC[ADJ] = 2 mg/m^3 × 6h/24h × 5d/7d = 0.36 mg/m^3

Keskin et al. 2009). No chronic or carcinogenicity animal studies on perineal exposure of talc were located in the literature.

A single injection of talc (in saline) into the bursa around the ovaries of rats showed foreign-body granulomas with confirmation of the presence of talc (Hamilton et al. 1984). Daily perineal or intravaginal application of talc (in saline) to rats for 3 months produced evidence of foreign-body reaction and infections; in addition, an increase in the number of inflammatory cells were found in all genital tissues. While no cancer or pre-cancer effects were observed, Keskin and colleagues (2009) noted that the study duration may have been too short to note these types of effects.

Human studies

Several meta-analyses of available epidemiological data have been published; some very recently (Huncharek et al. 2003; Langseth et al. 2008; Terry et al. 2013; Berge et al. 2018; Penninkilampi and Eslick 2018; Taher et al. 2018). These studies have consistently reported a positive association with ovarian cancer and perineal talc exposure. Taher and colleagues (2018) identified 27 studies (24 case-control and 3 cohort) for a meta-analysis; ever versus never perineal use of talc and the risk of ovarian cancer resulted in a statistically significant pooled odds ratio (OR) of 1.28 (see Table 6-1). Other published meta-analyses have demonstrated similar results, with ORs ranging from 1.22 to 1.35 (Huncharek et al. 2003; Langseth et al. 2008; Terry et al. 2013; Berge et al. 2018; Penninkilampi and Eslick 2018).

Table 6-1. Available human epidemiological studies investigating the association of perineal use of talc and ovarian cancer (Taher et al. 2018, in preparation)

Study type	Total sample size (no. of cases)	Study conclusion	OR [95% CI]	Reference
Case-control	686 (235)	Possible association in subgroup	Not included	Booth et al. 1989
Case-control	1014 (450)	Positive association	1.42 [1.08, 1.87]	Chang and Risch 1997
Case-control	336 (112)	Positive association in subgroup	Not included	Chen et al. 1992
Case-control	735 (313)	Positive association	1.60 [1.10, 2.33]	Cook et al. 1997
Case-control	430 (215)	Positive association	1.92 [1.27, 2.90]	Cramer et al. 1982
Case-control	4141 (2041)	Positive association	1.32 [1.15, 1.51]	Cramer et al. 2016
Case-control	3187 (1385)	Positive association	1.36 [1.14, 1.62]	Gates et al. 2008

Study type	Total sample size (no. of cases)	Study conclusion	OR [95% CI]	Reference
Case-control	305 (153)	No association	2.49 [0.94, 6.60]	Godard et al. 1998
Case-control	1684 (824)	Positive association	1.30 [1.10, 1.54]	Green et al. 1997
Case-control	274 (116)	No association	1.10 [0.70, 1.73]	Harlow and Weiss 1989
Case-control	474 (235)	Positive association in subgroup	1.50 [1.00, 2.25]	Harlow et al. 1992
Case-control	306 (135)	No association	0.70 [0.40, 1.22]	Hartge et al. 1983
Case-control	2704 (902)	Positive association	1.40 [1.16, 1.69]	Kurta et al. 2012
Case-control	225 (46)	No association	1.15 [0.41, 3.23]	Langseth and Kjaerheim 2004
Case-control	3085 (1576)	Positive association in subgroup	1.17 [1.01, 1.36]	Merritt et al. 2008
Case-control	1354 (249)	Positive association in subgroup	1.37 [1.02, 1.84]	Mills et al. 2004
Case-control	2143 (1086)	No association	1.06 [0.85, 1.32]	Moorman et al. 2009
Case-control	2134 (767)	Positive association in subgroup	1.50 [1.10, 2.05]	Ness et al. 2000
Case-control	123 (77)	Possible association	1.00 [0.20, 5.00]	Rosenblatt et al. 1992
Case-control	2125 (812)	Possible association	1.27 [0.97, 1.66]	Rosenblatt et al. 2011
Case-control	1329 (584)	Positive association	1.44 [1.11, 1.87]	Schildkraut et al. 2016
Case-control	389 (189)	No association	1.05 [0.28, 3.94]	Tzonou et al. 1993
Case-control	727 (188)	Possible association	1.45 [0.81, 2.60]	Whittemore et al. 1988
Case-control	1155 (462)	No association	1.00 [0.80, 1.25]	Wong et al. 1999
Case-control	1297 (609)	Positive association	1.53 [1.13, 2.07]	Wu et al. 2009
Case-control	4092 (1701)	Positive association in	1.46 [1.27, 1.68]	Wu et al. 2015

Study type	Total sample size (no. of cases)	Study conclusion	OR [95% CI]	Reference
		subgroup		
Cohort	108870 (797)	Possible association in subgroup	Not included	Gates et al. 2010
Cohort	78630 (307)	Possible association in subgroup	1.09 [0.86, 1.38]	Gertig et al. 2000
Cohort	41654 (154)	No association	0.73 [0.44, 1.21]	Gonzalez et al. 2016
Cohort	61285 (429)	No association	1.12 [0.92, 1.36]	Houghton et al. 2014

Abbreviation: CI, confidence interval.

Mode of action

The etiology of most ovarian tumours, in general, has not been well established. There are a number of different tumour types with characteristic histologic features, distinctive molecular signatures, and disease trajectories. Moreover, these tumours are heterogeneous, and they can arise from different tissues of the female reproductive tract, including the fallopian tube epithelium (National Academy of Sciences, Engineering, and Medicine 2016).

With respect to talc specifically, local chronic irritation leading to an inflammatory response is one possible mechanism of tumour progression that is frequently hypothesized (Muscat and Huncharek 2008; Penninkilampi and Eslick 2018; Taher et al. 2018). It is known that persistent indications of inflammation (including C-reactive protein, tumour necrosis factor, and other inflammatory markers) are detected in the blood of women prior to a diagnosis of ovarian tumours (Trabert et al. 2014). Increases in the number of inflammatory cells were found in all genital tissues of rats intravaginally exposed to talc for 3 months (Keskin et al. 2009). There is support for an association of inflammation and increased risk of ovarian cancer (National Academy of Sciences, Engineering and Medicine 2016; Rasmussen et al. 2017).

Talc particles were detected in the ovaries of rats that received intrauterine instillations of talc, and to a lesser extent in those that were dosed intravaginally with talc (Henderson et al. 1986). No translocation of talc into the ovaries was detected after single or multiple intravaginal applications of talc to rabbits (Phillips et al. 1978) or to monkeys (Wehner et al. 1986).

Talc particles were identified in 10 of 13 human ovarian tumours but were also found in 5 of 12 “normal” ovarian tissues removed from patients with breast cancer (Henderson et al. 1971). Ovaries from 24 patients undergoing incidental oophorectomy were examined; 12 women reported frequent perineal talc use, and the other 12 women were

non-users. Talc particles were detected in all 24 cases (both ever- and non-users) (Heller et al. 1996b). Wehner (2002) attributed the talc in the never users to (a) possible sample contamination, because some studies using negative controls resulted in particle counts similar to the test sample; and/or (b) possible false positives due to the use of a single radioactive tracer. To explain why talc is present in the never users, Heller and colleagues (1996b) hypothesized that talc use during diapering could contribute to the ovarian particle burden.

Translocation of other inert particles, similar in size to talc, has also been studied. A study in monkeys did not show any translocation of carbon black particles when a suspension was placed in the vaginal posterior fornix (Wehner et al. 1985). However, retrograde migration was detected when rabbits were administered a lubricant powder intravaginally (Edelstam et al. 1997). Other authors have noted similar transportation of particles to the upper genital tract (Egli and Newton 1961; De Boer 1972; Venter and Iturralde 1979). There are also some indications that particles can migrate from the vagina to the upper reproductive tract in humans (Egli and Newton 1961; Venter and Iturralde 1979; Heller et al. 1996a,b), and perineal exposure to talc has also been associated with a presence of talc in the lymph nodes and ovaries of women diagnosed with ovarian cancer (Heller et al. 1996a,b; Cramer et al. 2007).

Another possible mode of action that is hypothesized in the scientific literature is immune-mediated. It has been suggested that talc particles need not reach the ovaries but only need to reach the lower genital tract where talc could trigger changes (such as the production of heat shock proteins and/or decreased levels of antibodies) that could contribute to ovarian cancer (Cramer et al. 2005; Muscat et al. 2005). Human mucin 1 (MUC1) is expressed in high levels by ovarian cancer. Mucins are proteins involved in the formation of mucous barriers on epithelial surfaces (Gendler and Spicer 1995). Anti-MUC1 antibodies may have a protective effect; patients generate immunity against MUC1 produced by their tumours (Cramer et al. 2005). The Cramer et al. (2005) study used an enzyme-linked immunosorbent assay to measure anti-MUC1 antibody in women (controls; n = 721) to determine the factors that predict the presence of antibodies. It was found that the use of talc in the perineal area was associated with significantly decreased levels of antibodies to MUC1 (Cramer et al. 2005).

The most recent meta-analysis (Taher et al. 2018) employed the Hill criteria (Hill 1965) to assess the epidemiological evidence of a causal relationship. The Hill considerations are a set of factors (i.e., strength, consistency, specificity, temporality, biological gradient, biological plausibility, and coherence). These considerations form a framework for evaluating evidence in humans to help determine whether observed associations are causal (Hill 1965; Cogliano et al. 2004; US EPA 2005; Health Canada 2011; Fedak et al. 2015). Each factor, as reported in Taher et al. (2018), is elaborated upon below.

Strength: Of the 30 epidemiological studies examined by Taher et al. (2018), 15 case-control studies reported a positive association with statistical significance; 6 of these 15 had an OR of 1.5 or greater. Similarly, Penninkilampi and Eslick (2018) and Berge and colleagues (2018) each assessed 27 epidemiological studies and respectively

determined 14 and 13 case-control studies as reporting a positive association with statistical significance. In both cases, 5 of these studies had an OR of 1.5 or greater. Terry and colleagues (2013) only pooled 8 case-control studies; 5 of the 8 (63%) had a statistically significant positive association.

The individual cohort studies did not show a statistically significant association between perineal talc use and ovarian cancer (Berge et al 2018; Penninkilampi and Eslick 2018; Taher et al 2018). However, there was a positive association, with statistical significance, specific to invasive serous-type ovarian cancer in the cohort studies (OR = 1.25) (Penninkilampi and Eslick 2018). Given the long latency for ovarian cancer, the follow-up periods may not have been sufficient to capture all the cases for the individual cohort studies. Also, given the rarity of ovarian cancer, many of the available human studies may not be sufficiently powered to detect a low OR. Sample sizes were not large enough to detect a 20 to 30 % increase in risk; a group of over 200 000 women would need to be followed for over 10 years in order to detect a 20% (above background) increased risk with statistical significance (Narod 2016). With larger sample sizes, more individual studies may have demonstrated stronger associations.

Consistency: Several meta-analyses conducted over the past 15 years calculated similar ORs and resulted in similar conclusions; that there is a small yet consistent and statistically significant increased risk for ovarian cancer with perineal talc use (Huncharek et al. 2003; Langseth et al. 2008; Terry et al. 2013; Berge et al. 2018; Penninkilampi and Eslick 2018; Taher et al 2018). The epidemiological studies examined in these meta-analyses were conducted over different periods in time (across more than four decades), among different ethnicities, and spanned many geographical areas worldwide (Taher et al. 2018).

Specificity: Although there are many other risk factors for ovarian cancer (e.g., increased age, family history of cancer, obesity, nulliparity) (National Academy of Sciences, Engineering, and Medicine 2016), perineal talc exposure is specifically associated with cancer of the ovary and not other organs (Taher et al. 2018).

Temporality: In all case-control studies reporting positive outcomes, the participants recalled that exposure to talc preceded the reported outcome. However, in the cohort studies (reporting a lack of positive association), it is not known whether the follow-up period was adequate to detect a potential association between perineal talc exposure and ovarian cancer (Taher et al. 2018).

Biological gradient: There is a lack of an available exposure-effect relationship in the human epidemiological data. Many of the studies only assessed a single-dose level (ever versus never users). Furthermore, data with respect to the types of powder used by subjects or the amounts applied were not presented, and therefore a relationship between the concentration/dose of talc in the powder and the incidence of ovarian cancer could not be investigated. Taher and colleagues (2018) isolated seven studies that provided some evidence of increased risk of ovarian cancer with increasing perineal applications of talc; however, none demonstrated both a clear dose-response

trend and statistical significance (Whittemore et al. 1988; Harlow et al. 1992; Mills et al. 2004; Wu et al. 2009; Rosenblatt et al. 2011; Cramer et al. 2016; Schildkraut et al. 2016).

Biological plausibility: Particles of talc are hypothesized to migrate into the pelvis and ovarian tissue, causing irritation and inflammation. The presence of talc in the ovaries has been documented (Heller et al. 1996b). This evidence of retrograde transport supports the biologic plausibility of the association between perineal talc application and ovarian exposure; however, the specific mechanism(s) and cascade of molecular events by which talc might cause ovarian cancer have not been identified (Taher et al. 2018).

Coherence: Multiple case-control studies reported a lower risk of ovarian cancer in women who underwent pelvic surgery or tubal ligation (which disrupts the pathway and movement of talc from the lower to the upper genital tract) and suppressed ovulation (as cited by Taher et al. 2018: Cramer et al. 1982, 2016; Whittemore et al. 1988; Rosenblatt et al. 1992; Green et al. 1997; Wong et al. 1999; Mills et al. 2004). As noted in Penninkilampi and Eslick (2018), the main reductions in cancer incidence with tubal ligation were for serous and endometrial tumour types but not for mucinous or clear-cell tumours. Thus, tubal ligation is only effective in reducing the incidence of the same tumour types noted to be associated with perineal talc use.

The most recent meta-analysis detailed above (Taher et al. 2018), and consistent with the Hill criteria, suggests a small but consistent statistically significant positive association between ovarian cancer and perineal exposure to talc. Further, available data are indicative of a causal effect. A clear point of departure could not be derived from the available literature; consequently, hazard characterization is qualitative in nature.

6.2 Exposure assessment

This exposure assessment focuses on routes of exposure where critical effects have been identified; namely, non-cancer lung effects following inhalation of insoluble respirable particles of talc, and an association with ovarian cancer following perineal exposure to talc.

6.2.1 Environmental media, food and drinking water

Talc is a naturally occurring mineral, and there are several deposits in Canada (Kogel et al. 2006). Currently, there is one operating open-pit mine and concentrator along with an operating mill (MAC 2016); however, no talc concentration data in ambient air or around open-pit talc mines and processing facilities have been reported. Although particulate matter (PM) information for inhalable and respirable particles is available in the vicinity of these facilities (NPRI 2018), these data were not used in the exposure assessment as PM released from facilities is expected to contain a mixture of substances, hence the concentration would not reflect talc exposure from this source. However, given the

limited number of industrial and commercial sites producing and processing talc in Canada, talc exposure from ambient air is not expected to be significant.

Talc is insoluble in water (Table 3-1) and is expected to settle out during water treatment; exposure to the general population from drinking water is not expected.

There is potential for oral exposure resulting from the use of talc as a food additive; however, exposure from these uses is expected to be minimal (email from the Food Directorate, Health Canada, to the Existing Substances Risk Assessment Bureau, Health Canada, dated February 27, 2018; unreferenced). Exposure from the use of talc as a component in food packaging materials is expected to be negligible (email from the Food Directorate, Health Canada, to the Existing Substances Risk Assessment Bureau, Health Canada, dated February 27, 2018; unreferenced). Exposure from the oral route was not quantified because no critical health effects from the oral route of exposure have been identified. The JECFA has assigned an ADI of “not specified” for talc on the basis of low toxicity, and talc is “generally recognized as safe” as a food additive in the United States (JECFA 2006; U.S. FDA 2015).

6.2.2 Products available to consumers

Talc is present in approximately 8500 self-care products in Canada, including approximately 200 non-prescription drug products, approximately 2000 natural health products, and approximately 6500 cosmetic products. In addition, there are approximately 1300 prescription drugs containing talc. There is potential for oral exposure resulting from the use of self-care products and non-OTC drugs (including prescription, controlled substances, and ethical drugs) as a medicinal and non-medicinal ingredient containing talc. However, exposure from the oral route was not quantified as no critical health effects from the oral route of exposure have been identified.

There is the potential for dermal contact with talc from the use of self-care products. Systemic exposure resulting from dermal contact with talc is expected to be negligible as it is not expected that talc will be absorbed on the basis of its physical-chemical characteristics as an insoluble solid particle. In addition, a dermal health effect endpoint has not been identified for talc.

Notifications submitted under the *Cosmetic Regulations* to Health Canada for talc, the LNHPD (modified 2018), the Drug Product Database (DPD), voluntary information submitted to Environment and Climate Change Canada and Health Canada (ECCC, HC 2017), publicly available databases and websites (e.g., Household Products Database 1993-; CPCat 2014; CPID 2017), and material safety and technical datasheets were used to identify products where there is: (a) the potential for inhalation of insoluble respirable talc, and (b) the potential exposure to the perineal region. These products and associated exposures are presented below.

No inhalation or perineal exposures were identified with respect to the major commercial or industrial uses of talc in paper, plastics, ceramics, and putties.

Inhalation exposure

For inhalation exposure, potential exposures were focused on products that were formulated as loose powders and were available to consumers, which included approximately 400 self-care products (primarily cosmetics). Products formulated as pressed powders, which comprise the majority of cosmetics containing talc (approximately 4000 products) were not identified as a potential source of exposure of concern because the formation of a “dust cloud” available for inhalation is not expected during the use of these products. Available information of interest were self-care products marketed as cosmetics, NHPs, or non-prescription drugs that are intended for application to the body, face, feet, buttocks (babies), and hair (e.g., dry hair shampoo). Concentrations of talc range from less than 10 to 100 % in these types of products.

In order to determine if talc loose-powder self-care products contain respirable particles, Health Canada measured the particle size distribution of three products (one baby powder and two adult body powder products) containing high concentrations of talc (>90%) available in Canada (Rasmussen 2017). Using an Aerodynamic Particle Sizer, the particle size distribution for the three products ranged from < 0.5 µm to 8 µm, with median particle sizes ranging from 1.7 to 2 µm. Thus, all of the particles were within the inhalable range (< 10 µm), and the median particle size was within the respirable range (< 4 µm). Number concentrations measured using a scanning mobility particle sizer indicated that the proportion of nano-sized particles (<100 nm) was small (< 10 %) to negligible, depending on the product.

Several studies were conducted by the cosmetic industry in the 1970s to provide data required to assess the safety of talc powder products and generate air concentrations (Aylott et al. 1979; Russell et al. 1979). These studies demonstrated that during the use of face, baby, and adult powders, there are quantifiable concentrations of respirable talc particles available for inhalation exposure. In 1978, Aylott and colleagues determined mean respirable air concentrations of 0.48 to 1.9 mg/m^3 of talc (< 7 µm) over 5 minutes for loose face powder, adult dusting powder, baby dusting powder, and micronized adult dusting powder. That same year, concentrations of talc (< 10 µm) of 0.19 mg/m^3 and 2.03 mg/m^3, respectively, were determined near the infant breathing zone during a simulation of routine application of talcum powder during diapering, and in the breathing zone of adults during the application of talcum powder to their body (Russell et al. 1979). In both of these studies, the highest air concentrations were associated with the adult application of talcum powder to their bodies over infant diapering and application of loose facial powder. There are uncertainties with the calculated talc concentrations determined from these studies due to limitations in the collection and analysis of talc concentrations on the basis of the use of older equipment, older sampling methods, and older talc products.

In 2017, a study assessing the health risk from the use of cosmetic talc from historical products was published (Anderson et al. 2017). This study included examining historical talc products from the 1960s and 1970s to characterize airborne respirable dust concentrations during the use of these products. To quantify respirable talc concentrations in the breathing zone, Anderson and colleagues (2017) designed a study where 5 volunteers were asked to apply historical talc products as they typically would in a bathroom setting. Cyclone air sampling devices were attached to the breathing zone of each volunteer. Each exposure simulation consisted of 8 application events, at six-minute intervals, for a total sampling duration of 48 minutes. This study design ensured that the sample mass on the sampling filter was large enough for quantification and accuracy, but it was not expected that during the typical use of a talc body powder that individuals apply talc every six minutes over a 48-minute window. Average talc concentrations over the 48-minute exposure simulation were calculated using the total measured mass (from 8 applications over 48 minutes) and the air volume over the entire 48-minute sampling period. Respirable talc concentrations ranged from 0.26 to 5.03 mg/m^3, and the average was 1.46 mg/m^3. The average air concentration by subject ranged from 0.44 to 3.28 mg/m^3. Respirable talc concentrations were more variable between subjects than within subjects, suggesting that individual behaviour has a strong influence in airborne concentrations.

In 2018, Health Canada conducted a small study in order to measure the air concentrations of particles in the breathing zone of adult volunteer subjects while they were applying talc-containing self-care products (Rasmussen 2018). Continuous, direct-reading, personal breathing-zone monitors (positioned beside the nose) measured average particulate matter of aerodynamic diameter of 4 µm or less (PM_4) concentrations of 0.48 ± 0.18 mg/m^3 and 1.80 ± 0.82 mg/m^3 for volunteers applying body powder and loose face powder, respectively. Subjects repeated the application in triplicate. These average concentrations fall within the range of concentrations measured by Anderson and colleagues (2017). In this study, the application of loose face powder resulted in the highest average air concentration in the immediate vicinity of the nose.

Several exposure scenarios were derived to characterize inhalation exposure to talc particles from the use of self-care products; namely, the use of baby, body, face, and foot powders (loose formulations), and dry hair shampoo. Average air concentrations by subject from Anderson et al. 2017 were combined with the body and face replicates from Rasmussen 2018 to obtain an overall average air concentration of 1.36 ± 0.97 mg/m^3. This value was used to estimate adjusted air concentrations for self-care products based on the highest concentration of talc present in these products. The results are summarized in Table 6-2. The inputs for each of these scenarios are outlined in Appendix A.

Table 6-2. Inhalation exposure estimates to talc from self-care products available to consumers

Product type	Age group	Concentration in air per event (mg/m^3)[a]	Adjusted exposure concentration (mg/m^3)[b]
Baby powder 100% talc	Infant and Adult	1.36	0.0071
Body powder 100% talc	Adult	1.36	0.0047
Face powder 100% talc	Adult	1.36	0.0047
Foot powder 97% talc	Adult	1.32	0.0034
Dry hair shampoo 100% talc	Adult	1.36	0.0011

[a] Average measured air concentrations (Anderson et al. 2017, Rasmussen 2018) × the highest concentration of talc in product type.
[b] Refer to Appendix A for details.

Perineal exposure

Several types of self-care products have the potential to result in exposure to the perineal region. There are several baby and body powders (approximately 50 products) with concentrations of talc that range from 0.3 to 100 %. There has been a decline in popularity of the use of talc for feminine hygiene practices over time; of 6000 North American women, 19 % of women born between 1920 and 1940 reported applying talc directly to the perineal region, but only 3% of women born after 1975 reported the same (Narod 2016). Houghton and colleagues (2014) reported that in 2001, the proportion of U.S. women who were users of perineal talc was estimated at 40 %, down from 52 % during 1993 to 1998.

There is a small number of diaper or rash cream self-care products (less than 10) which contains low concentrations of talc as a non-medicinal ingredient (up to 0.5 %). Talc is permitted as a medicinal ingredient in diaper rash products at concentrations from 45 to 100 % (Health Canada 2007); however, there are no diaper rash products listed in the LNHPD containing talc as a medicinal ingredient (LNHPD [modified 2018]).

Additional self-care products that have the potential for perineal exposure (approximately 100 products) include antiperspirants and deodorants (e.g., genital antiperspirants), body wipes, bath bombs, and to a lesser extent (due to wash off or removal) other bath products (i.e., soap, shower gel) and products associated with hair removal (e.g., epilatory products). These products are formulated as gels, sprays, loose powders, and solid cakes, and range in concentration from less than 1% to 100% talc.

As indicated in Section 4, there is no evidence to suggest that talc is currently being used as a dry lubricant on condoms or medical examination gloves in Canada. At present, these are not considered to be sources of perineal exposure.

As a quantitative point of departure could not be derived from the available literature, perineal exposure from the use of self-care products was not quantified.

6.3 Characterization of risk to human health

Consistent with other international regulatory and advisory bodies (Danish EPA, U.S. EPA, MAK-Commission, U.S. FDA, and JECFA), no critical health effects were identified via the oral or dermal routes of exposure. As such, oral exposure to talc resulting from food intake and use of self-care products are not of concern.

Critical health effects have been identified following inhalation exposure to respirable talc particles. From the available toxicological studies, a NOAEC of 2 mg/m^3 from the NTP inhalation studies in mice and rats was identified in which non-cancer lung effects, with lung overload, were noted at the next highest concentration of 6 mg/m^3.

The average air concentration of talc following the use of a loose-powder self-care product (1.36 mg/m^3) provides a small margin of exposure (i.e., 1.5) to the NOAEC of 2 mg/m^3. However, the NOAEC is derived from a study with an exposure profile of 6 hours per day, 5 days per week, over 4 weeks, while the actual exposure scenarios from the use of self-care products are intermittent, occurring in minutes per day, daily, or weekly over many years. To address the differences in exposure between the NTP study and the actual use pattern, both the NOAEC and the talc air concentrations were adjusted to a continuous exposure scenario according to U.S. EPA guidance on inhalation risk assessment to more accurately characterize potential risk (U.S. EPA 1994, 2009). The NOAEC of 2 mg/m^3 is equivalent to an adjusted concentration of 0.36 mg/m^3, as noted in the Health Effects section. The NOAEC of 2 mg/m^3 was extracted from a 4-week inhalation study as a NOAEC for chronic exposure was not available. Episodic exposures from product use are expected to increase lung load due to the long alveolar clearance of talc. The adjusted air concentrations from the use of self-care products are presented in Table 6-3.

Table 6-3. Relevant exposure and hazard values for talc, and margins of exposure, for determination of risk

Exposure scenario	Adjusted air concentration, CA (mg/m^3)[a]	Adjusted critical-effect level (mg/m^3)	Critical health effect endpoint	MOE
Baby powder 100% talc	0.0071	NOAEC[adj]: 0.36	non-cancer lung effects	50

Body powder 100% talc	0.0047	NOAEC[adj]: 0.36	non-cancer lung effects	76
Face powder 100% talc	0.0047	NOAEC[adj]: 0.36	non-cancer lung effects	76
Foot powder 97% talc	0.0034	NOAEC[adj]: 0.36	non-cancer lung effects	106
Dry hair shampoo 100% talc	0.0011	NOAEC[adj]: 0.36	non-cancer lung effects	327

Abbreviations: adj, adjusted; CA, concentration in air per event; MOE, margin of exposure.
[a] From Anderson et al. (2017) and Rasmussen (2018), respectively, based on the highest concentration in products. For most of these product types, there is a wide range of talc concentrations (< 10 to 100 %).

The margins of exposure (MOEs) between the adjusted critical-effect level and the adjusted air concentrations range from 50 to 327 for self-care products. The MOEs for baby powder, body powder, face powder, and foot powder are considered potentially inadequate to account for uncertainties in the health effects (including a lack of a NOAEC from chronic studies) and exposure databases. The MOE for dry hair shampoo is considered adequate to address uncertainties in the health effects and exposure databases.

Based on available human data, ovarian cancer was also identified as a critical health effect for the perineal route of exposure to talc. There is the potential for perineal exposure to talc from the use of various self-care products (e.g., body powder, baby powder, diaper and rash creams, genital antiperspirants and deodorants, body wipes, bath bombs). As noted in the Health Effects section, a point of departure cannot be derived for this health effect. Data from published meta-analyses of epidemiological studies indicate a consistent and statistically significant positive association between perineal exposure to talc and ovarian cancer (Huncharek et al. 2003; Langseth et al. 2008; Terry et al. 2013; Berge et al. 2018; Penninkilampi and Eslick 2018; Taher et al. 2018). As noted by Narod (2016), “It is unlikely that the association between talc and ovarian cancer is due to confounding and so it is fair to say that if there is a statistically robust relationship between talc use and ovarian cancer it is likely to be causal.” Similarly, Penninkilampi and Eslick (2018) noted that “the confirmation of an association in cohort studies between perineal talc use and serous invasive ovarian cancer is suggestive of a causal association.” Taher and colleagues (2018) noted that “consistent with previous evaluations by the International Agency for Research on Cancer (2010), and more recent and subsequent evaluations by individual investigators (Penninkilampi and Eslick 2018; Berge et al. 2018; Terry et al. 2013), the present comprehensive evaluation of all currently available relevant data indicates that perineal exposure to talc powder is a possible cause of ovarian cancer in humans.”

The meta-analyses of the available human studies in the peer-reviewed literature indicate a consistent and statistically significant positive association between perineal exposure to talc and ovarian cancer. Further, available data are indicative of a causal effect. Given that there is the potential for perineal exposure to talc from the use of various self-care products, a potential concern for human health has been identified.

6.4 Uncertainties in evaluation of risk to human health

The inhalation of talc has been associated with a variety of non-cancerous lung effects, commonly termed talcosis. Dose-response data for lung effects in humans is, for the most part, lacking, and the use of animal data to quantify risk due to talc inhalation is considered appropriate. Despite the lack of exposure quantification, there are numerous case reports, as well as worker studies, that have identified non-cancer health effects from inhalation of talc powders. There is some uncertainty regarding the extrapolation of the NOAEC identified in animal models exposed for 6 hours per day for a short duration (4 weeks) to long-term episodic human exposures. The true NOAEC for chronic exposure is likely substantially lower than 2 mg/m^3.

Some self-care products, in particular, some face powders, may contain a cover or another mechanism that would reduce the potential for the generation of a particle or dust cloud, or that would reduce the concentration of the dust cloud during use of the product. There is uncertainty as to which products, and the proportion of products on the market, that incorporate these exposure-mitigation measures.

There are limitations with the human epidemiological data. Potential sources of bias include selection bias due to low response rates or from limiting subjects, and exposure misclassification due to recall bias (Taher et al. 2018). Muscat and Huncharek (2008) also proposed that symptoms of ovarian cancer prior to diagnosis may increase the perineal use of talc and bias the results. However, Narod (2016) and Berge and colleagues (2018) put less emphasis on recall bias. In studies where the exposure is simple (e.g., never versus ever use), recall bias is unlikely to be an important source of bias (Narod 2016). The positive association is strongest for the serous histologic type (Berge et al. 2018; Taher et al. 2018); findings that the association may vary by histologic type detracts from the hypothesis of report bias, as this type of bias would likely operate for all histologic types (Berge el al. 2018).

Ovarian cancer, in general, is not well understood (National Academy of Sciences, Engineering, and Medicine 2016), and a comparable animal model is not available. Health Canada has identified self-care products with the potential for perineal exposure (e.g., baby powder, body powders, diaper and rash creams, genital antiperspirants and deodorants, body wipes, bath bombs); however, there is no indication exactly how the products are being used, the extent to which they would contribute to perineal exposure, and with what frequency and amount.

Talc use during diapering is a confounder that was not adequately accounted for in the epidemiological studies. It has not been determined whether the internal female genital

tract is exposed to talc dusts during infancy (Muscat and Huncharek 2008). As well, not all the available human studies are clear as to the formulations used for perineal applications. It is possible that the identified cancer incidences are specific to loose-powder formulations; however, there is inadequate information to attribute the cancer incidences to other formulation types (e.g., creams).

7. Conclusion

Considering all available lines of evidence presented in this draft screening assessment, there is low risk of harm to the environment from talc. It is proposed to conclude that talc does not meet the criteria under paragraphs 64(*a*) or (*b*) of CEPA as it is not entering the environment in a quantity or concentration or under conditions that have or may have an immediate or long-term harmful effect on the environment or its biological diversity or that constitute or may constitute a danger to the environment on which life depends.

On the basis of the information presented in this draft screening assessment, it is proposed to conclude that talc meets the criteria under paragraph 64(*c*) of CEPA as it is entering or may enter the environment in a quantity or concentration or under conditions that constitute or may constitute a danger in Canada to human life or health.

It is therefore proposed to conclude that talc meets one of the criteria set out in section 64 of CEPA.

Talc is proposed to meet the persistence criteria but not the bioaccumulation criteria as set out in the *Persistence and Bioaccumulation Regulations* of CEPA.

References

Akira M, Kozuka T, Yamamoto S, Sakatani M, Morinaga K. 2007. Inhalational talc pneumoconiosis: radiographic and CT findings in 14 patients. Am J Roentgenol. 188(2):326-333.

Anderson EL, Sheehan PJ, Kalmes RM, Griffin JR. 2017. Assessment of Health Risk from Historical Use of Cosmetic Talcum Powder. Risk Anal. 37(5):918-928.

Aylott RI, Byrne GA, Middleton JD, Roberts. 1979. Normal use levels of respirable cosmetic talc: preliminary study. Int J Cosmet Sci. 1:177-186.

Berge W, Mundt K, Luu H, Boffetta P. 2018. Genital use of talc and risk of ovarian cancer: a meta-analysis. Eur J Cancer Prev. 27(3):248-257.

Booth M, Beral V, Smith P. 1989. Risk factors for ovarian cancer: a case-control study. Br J Cancer. 60(4):592-598.

Canada. 1999. *Canadian Environmental Protection Act, 1999*. S.C. 1999, c.33. Canada Gazette Part III, vol. 22, no. 3.

Carr CJ. 1995. Talc: Consumer Uses and Health Perspectives. Proceedings of a workshop. Bethesda, Maryland, January 31–February 1, 1994. Regul Toxicol Pharmacol. 21(2):211-215.

Cevc G. 1997. Drug delivery across the skin. Expert Opin Inv Drug. 6(12):1887-1937.

Chang S, Risch HA. 1997. Perineal talc exposure and risk of ovarian carcinoma. Cancer. 79(12):2396-2401.

Chang CJ, Tu YK, Chen PC, and Yang HY. 2017. Occupational exposure to talc increases the risk of lung cancer: A meta-analysis of occupational cohort studies. Can Respir J. 2017:1-12.

ChemIDplus [database]. 1993-. Bethesda (MD): U.S. National Library of Medicine. [updated 2017 April 11; accessed 2017 May 26].

Chen Y, Wu PC, Lang JH, Ge WJ, Hartge P, Brinton LA. 1992. Risk factors for epithelial ovarian cancer in Beijing, China. Int J Epidemiol. 21(1):23-29.

[CIMT] Canadian International Merchandise Trade Database [database]. 2017. Ottawa (ON): Government of Canada. [accessed 2017 October].

[CIR] Cosmetic Ingredient Review Expert Panel. 2013. Safety Assessment of Talc as Used in Cosmetics. Final Report [PDF]. Washington (DC): Cosmetic Ingredient Review. [accessed 2017 November].

Cogliano VJ, Baan RA, Straif K, Grosse Y, Secretan MB, Ghissassi FE, Kleihues P. 2004. The science and practice of carcinogen identification and evaluation. Environ Health Perspect. 112(13):1269-1274.

Cook LS, Kamb ML, Weiss NS. 1997. Perineal powder exposure and the risk of ovarian cancer. Am J Epidemiol. 145(5):459-465.

[CPCat] Chemical and Product Categories [database]. 2014. Ver. 04. Washington (D.C.): U.S. Environmental Protection Agency. [updated 2014 May 21; accessed 2014 Nov 21]. [Database described

in Dionisio KL, Frame AM, Goldsmith MR, Wambaugh JF, Liddell A, Cathey T, Smith D, Vail J, Ernstoff AS, Fantke P, et al. 2015. Exploring consumer exposure pathways and patterns of use for chemicals in the environment. Toxicol Rep. (2):228-237.].

[CPID] Consumer Product Information Database [database]. 2017. McLean (VA): DeLima Associates. [accessed 2017 Nov 21].

Cramer DW, Vitonis AF, Terry KL, Welch WR, Titus LJ. 2016. The Association Between Talc Use and Ovarian Cancer: A Retrospective Case-Control Study in Two US States. Epidemiology. 27(3):334-346.

Cramer DW, Welch WR, Berkowitz RS and Godleski JJ. 2007. Presence of talc in pelvic lymph nodes of a woman with ovarian cancer and long term genital exposure to cosmetic talc. Obstet Gynecol. 110(2 Pt 2):498-501.

Cramer DW, Titus-Ernstoff L, McKolanis JR, Welch WR, Vitonis AF, Berkowitz RS, Finn OJ. 2005. Conditions associated with antibodies against the tumor-associated antigen MUC1 and their relationship to risk for ovarian cancer. Cancer Epidemiol Biomarkers Prev. 14(5):1125-1131.

Cramer DW, Welch WR, Scully RE, Wojciechowski CA. 1982. Ovarian cancer and talc: a case-control study. Cancer. 50(2):372-376.

Cruthirds TP, Cole FH, Paul RN. 1977. Pulmonary talcosis as a result of massive aspiration of baby powder. South Med J. 70(5):626-628.

[CTFA] Cosmetic, Toiletry and Fragrance Association. 1983. Summary for the Results of Surveys of the amount and Frequency of use of cosmetic products by Women. Report Prepared by Pitkin B, Rodericks JV, Turnbull D. Washington (DC): CTFA Inc.

[Danish EPA] Danish Environmental Protection Agency. 2016. Evaluation of health hazards by exposure to talcum, cosmetic grade (non-fibrous) and proposal of a health-based quality criterion for ambient air [PDF]. Denmark: Danish Environmental Protection Agency. ISBN: 978-87-93529-23-6.

De Boer CH. 1972. Transport of particulate matter through the human female genital tract. J Reprod Fertil. 28(2):295-297.

Douglas A, Karov J, Daka J, Hinberg I. 1998. Detection and Quantitation of Talc on Latex Condoms. Contraception. 58(3):153-155.

[DPD] Drug Product Database [database]. [modified 2018 June 12]. Ottawa (ON): Government of Canada. [accessed 2018 Aug 15].

Environment Canada. 2013. DSL Inventory Update data collected under the *Canadian Environmental Protection Act, 1999*, section 71: *Notice with respect to certain substances on the Domestic Substances List.* Data prepared by: Environment Canada, Health Canada; Existing Substances Program.

[ECCC] Environment and Climate Change Canada. 2018. Science approach document: ecological risk classification of inorganic substances. Ottawa (ON): Government of Canada.

[ECCC, HC] Environment and Climate Change Canada, Health Canada. 2017. Targeted information gathering for screening assessments under the Chemicals Management Plan (February to July 2017). Data prepared by: ECCC, Health Canada; Existing Substances Program.

[ECCC, HC] Environment and Climate Change Canada, Health Canada. [modified 2017 Mar 12]. Categorization of chemical substances. Ottawa (ON): Government of Canada. [accessed 2018 Aug 30].

Edelstam GAB, Sjösten ACE, Ellis, H. 1997. Retrograde migration of starch in the genital tract of rabbits. Inflammation. 21(5):489-499.

Egli GE, Newton M. 1961. The transport of carbon particles in the human female reproductive tract. Fertil Steril. 12:151-155.

[EU] Commission of the European Communities. [modified 2001 Oct 1]. Report from the Commission on Dietary Food Additive Intake in the European Union. Brussels (BE): Commission of the European Communities.

[EuroTalc] Scientific Association of European Talc Producers. 2017. "What is talc?" Brussels (BE): Eurotalc. [accessed 2017 May 29]

[FAO] Food and Agriculture Organization of the United Nations. 2006. Combined Compendium of Food Additives Specifications: Sixty-first meeting of the Joint FAO/WHO Expert Committee on Food Additives. FAO Food and Nutrition Paper 52.

Fedak KM, Bernal A. Capshaw ZA, Gross S. 2015. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. Emerg Themes Epidemiol. 12:14.

Feigin DS.1986.Talc: understanding its manifestations in the chest. Am J Roentgenol. 146(2):295-301.

Fiume MM, Boyer I, Bergfeld WG, Belsito DV, Hill RA, Klaassen CD, Liebler DC, Marks Jr JG, Shank RC, Slaga TH, Snyder PW, Anderson FA. 2015. Safety Assessment of Talc Used in Cosmetics. Int J Toxicol. 34(1 suppl):66S-129S.

Frank C, Jorge L. 2011. An uncommon hazard: Pulmonary talcosis as a result of recurrent aspiration of baby powder. Respir Med CME. 4(3):109-111.

Ficheux AS, Wesolek N, Chevillotte G, Roudot AC. 2015. Consumption of cosmetic products by the French population. First part: Frequency data. Food Chem Toxicol. 78:159-169.

Frosch PJ, Kligman AM. 1976. The chamber-scarification test for irritancy. Contact Derm. 2:314-324.

Gates MA, Tworoger SS, Terry KL, Titus-Ernstoff L, Rosner B, De Vivo I, Cramer DW, Hankinson SE. 2008. Talc use, variants of the GSTM1, GSTT1, and NAT2 genes, and risk of epithelial ovarian cancer. Cancer Epidemiol Biomarkers Prev. 17(9):2436-2444.

Gates MA, Rosner BA, Hecht JL, Tworoger SS. 2010. Risk factors for epithelial ovarian cancer by histologic subtype. Am J Epidemiol. 171(1):45-53.

Gendler SJ, Spicer AP. 1995. Epithelial mucin genes. Annu Rev Physiol. 57:607-634.

Gertig DM, Hunter DJ, Cramer DW, Colditz GA, Speizer FE, Willett WC, Hankinson SE. 2000. Prospective study of talc use and ovarian cancer. J Natl Cancer Inst. 92(3):249-252.

Gibbs AE, Pooley FD, Griffiths DM, Mitha R, Craighead JE, Ruttner JR. 1992. Talc pneumoconiosis: a pathologic and mineralogic study. Hum Pathol. 23(12):1344-1354.

Gibel W, Lohs K, Horn KH, Wildner GP, Hoffmann F. 1976. Experimental study on cancerogenic activity of asbestos filters. Arch Geschwulstforsch. 46:437-442.

Godard B, Foulkes WD, Provencher D, Brunet JS, Tonin PN, Mes-Masson AM, Narod SA, Ghadirian P. 1998. Risk factors for familial and sporadic ovarian cancer among French Canadians: a case-control study. Am J Obstet Gynecol. 179(2):403-410.

Gonzalez NL, O'Brien KM, D'Aloisio AA, Sandler DP, Weinberg CR. 2016. Douching, talc use, and risk of ovarian cancer. Epidemiology. 27(6):797-802.

Gould SR, and Barnardo DE. 1972. Respiratory distress after talc inhalation. Brit J Dis Chest. 66:230-233.

Green A, Purdie D, Bain C, Siskind V, Russell P, Quinn M, Ward B. 1997. Tubal sterilisation, hysterectomy and decreased risk of ovarian cancer. Survey of Women's Health Study Group. Int Cancer. 71(6):948-951.

Gysbrechts C, Michiels E, Verbeken E, Verschakelen J, Dinsdale D, Nemery B, Demedts M. 1998. Interstitial lung disease more than 40 years after a 5 year occupational exposure to talc. Eur Respir J. 11(6):1412-1415.

Hamilton TC, Fox H, Buckley CH, Henderson WJ, Griffiths K. 1984. Effects of talc on the rat ovary. Br J Exp Pathol. 65(1):101-106.

Harlow BL, Weiss NS. 1989. A case-control study of borderline ovarian tumors: the influence of perineal exposure to talc. Am J Epidemiol. 130(2):390-394.

Harlow BL, Cramer DW, Bell DA, Welch WR. 1992. Perineal exposure to talc and ovarian cancer risk. Obstet Gynecol. 80(1):19-26.

Hartge P, Hoover R, Lesher LP, McGowan L. 1983. Talc and ovarian cancer. J Am Med Assoc. 250(14):1844.

Health Canada. 2007. Diaper rash products [PDF]. Ottawa (ON): Government of Canada.

Health Canada. 2010. PMRA list of formulants [PDF]. Ottawa (ON): Government of Canada.

Health Canada. 2011. Weight of evidence: general principles and current applications at Health Canada. November 2011. Unpublished report. Prepared by the Task Force on Scientific Risk Assessment's Weight of Evidence Working Group.

Health Canada. 2015. Natural Health Product Traditional Chinese Medicine Ingredients (TCMI). Ottawa (ON): Government of Canada.

Health Canada. [modified 2017 May 3]. List of permitted food additives. Ottawa (ON): Government of Canada. [accessed 2017 May 29].

Health Canada. [modified 2018 Jun 14]. Cosmetic ingredient hotlist: list of ingredients that are prohibited for use in cosmetic products. Ottawa (ON): Government of Canada. [accessed 2018 Aug 30].

Heller DS, Gordon RE, Westhoff C, Gerber S. 1996a. Asbestos exposure and ovarian fiber burden. Am J Ind Med. 29:435-439.

Heller DS, Westhoff C, Gordon RE, Katz N. 1996b. The relationship between perineal cosmetic talc usage and ovarian talc particle burden. Am J Obstet Gynecol. 174(5):1507-1510.

Henderson WJ, Joslin CAF, Griffiths K, Turnbull AC. 1971. Talc and carcinoma of the ovary and cervix. BJOG: Int J Obstet Gynaecol. 78(3):266-272.

Henderson WJ, Hamilton TC, Baylis MS, Pierrepoint CG, Griffiths K. 1986. The demonstration of the migration of talc from the vagina and posterior uterus to the ovary in the rat. Environ Res. 40(2):247-250.

Hill AB. 1965. The environment and disease: association or causation? Proc R Soc Med. 58:295-300.

Hollinger MA. 1990. Pulmonary toxicity of inhaled and intravenous talc. Toxicol Lett. 52(2):121-127; discussion 117-119.

Houghton SC, Reeves KW, Hankinson SE, Crawford L, Lane D, Wactawski-Wende J, Thomson CA, Ockene JK, Sturgeon SR. 2014. Perineal powder use and risk of ovarian cancer. J Natl Cancer Inst. 106(9).

Household Products Database [database]. 1993-. Bethesda (MD): National Library of Medicine (US). [updated 2016 September; accessed 2017 June 19].

[HSDB] Hazardous Substances Data Bank [database]. 2005. CAS RN 14807-96-6. Bethesda (MD): National Library of Medicine (US). [complete update 2005 May 2; accessed 2017 Nov 21].

Huncharek M, Geschwind JF, Kupelnick B. 2003. Perineal application of cosmetic talc and risk of invasive epithelial ovarian cancer: a meta-analysis of 11,933 subjects from sixteen observational studies. Anticancer Research 23(2C):1955-1960.

[IARC] International Agency for Research on Cancer. 1987. Talc not containing asbestiform fibres (group 3). Talc containing asbestiform fibres (group 1). Summaries & Evaluations. Suppl 7:349.

[IARC] International Agency for Research on Cancer. 2010. Carbon Black, Titanium Dioxide, and Talc, IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. 93:277-413.

[ISO] International Organization for Standardization. 2015. ISO 4074: 2015 Natural rubber latex male condoms – Requirements and test methods. Geneva (CH): International Organization for Standardization.

[JECFA] Joint FAO/WHO Expert Committee on Food Additives. 2006. Compendium of Food Additive Specifications. FAO JECFA Monograph 1.

Keskin N, Teksen YA, Ongun EG, Ozay Y, Saygili H. 2009. Does long-term talc exposure have a carcinogenic effect on the female genital system of rats? An experimental pilot study. Arch Gynecol. 280(6):925-931.

Kogel JE, Trivedi NC, Barker JM, Krukowski ST, eds. 2006. Industrial Minerals and Rocks. 7th ed. Littleton (CO): Society for Mining, Metallurgy, and Exploration, Inc.

Kurta ML, Moysich KB, Weissfeld JL, Youk AO, Bunker CH, Edwards RP, Modugno F, Ness RB, Diergaarde B. 2012. Use of fertility drugs and risk of ovarian cancer: results from a U.S.-based case-control study. Cancer Epidemiol Biomarkers Prev. 21(8):1282-1292.

Langseth H, Kjærheim K. 2004. Ovarian cancer and occupational exposure among pulp and paper employees in Norway. Scand J Work Environ Health. 30(5):356-361.

Langseth H, Hankinson SE, Siemiatycki J, Weiderpasse E. 2008. Perineal use of talc and risk of ovarian cancer. J Epidemiol Community Health. 62(4):358-360.

[LNHPD] Licensed Natural Health Products Database [database]. [modified 2018 Feb 6]. Ottawa (ON): Government of Canada. [accessed 2018 Aug 14].

Lundberg M, Wrangsjo K, Johansson SGO. 1997. Latex allergy from glove powder – an unintended risk with the switch from talc to cornstarch. Allergy 52:1222-1228.

[MAC] Mining Association of Canada. 2016. Facts and Figures of the Canadian Mining Industry F&F 2016 [PDF]. [accessed 2017 Nov 21].

[MAK-Commission] The MAK Collection for Occupational Health and Safety. 2012. Talc (without asbestos fibres) (respirable fraction). Weinheim (DE): Wiley-VCH Verlag GmbH & Co. KGaA. The MAK-collection Part I: MAK Value Documentations, Vol. 22. 226-279.

Marchiori E, Lourenço S, Gasparetto TD, Zanetti G, Mano CM, Nobre LF. 2010. Pulmonary talcosis: imaging findings. Lung. 188(2):165-171.

Merritt MA, Nagle CM, Webb PM, Bowtell D, Chenevix-Trench G, Green A, DeFazio A, Gertig D, Traficante N, Moore S, et al. 2008. Talcum powder, chronic pelvic inflammation and NSAIDs in relation to risk of epithelial ovarian cancer. Int J Cancer. 122(1):170-176.

Mills PK, Riordan DG, Cress RD, Young HA. 2004. Perineal talc exposure and epithelial ovarian cancer risk in the Central Valley of California. Int J Cancer. 112(3):458-464.

Moorman PG, Palmieri RT, Akushevich L, Berchuck A, Schildkraut JM. 2009. Ovarian cancer risk factors in African-American and white women. Am J Epidemiol. 170(5):598-606.

Muscat JE, Huncharek, MS. 2008. Perineal talc use and ovarian cancer: a critical review. Eur J Cancer Prev. 17(2):139-146.

Muscat J, Huncharek M, Cramer DW. 2005. Talc and anti-MUC1 antibodies. Cancer Epidemiol Biomarkers Prev. 14(11 Pt. 1):2679.

Narod SA. 2016. Talc and ovarian cancer. Gynecol Oncol. 141:410-412.

National Academy of Sciences, Engineering, and Medicine. 2016. Ovarian cancers: evolving paradigms in research and care. Washington (D.C.): National Academy Press.

Ness RB, Grisso JA, Cottreau C, Klapper J, Vergona R, Wheeler JE, Morgan M, Schlesselman JJ. 2000. Factors related to inflammation of the ovarian epithelium and risk of ovarian cancer. Epidemiology 11(2):111-117.

[NHPID] Natural Health Products Ingredients Database [database]. [modified 2018 July 6]. Ottawa (ON): Government of Canada. [accessed 2018 Aug 14].

[NIOSH] National Institute for Occupational Safety and Health (US). 2014. Talc (silica and fibre free). International Chemical Safety Card (ICSC). Atlanta (GA): Centre for Disease Control. ICSC # 0329. [accessed 2018 Mar].

[NPRI] National Pollutant Release Inventory. 2018. NPRI Datasets: Substance: PM10 - Particulate Matter <= 10 Microns, Company/Facility information: Imerys Talc Canada Inc. (2017). Ottawa (ON): Government of Canada. Search results for PM_{10} at Imerys Talc Canada Inc. [updated 2018 June 14].

[NTP] National Toxicology Program. 1993. NTP technical report on the toxicology and carcinogenesis studies of talc (CAS NO. 14807-96-6) in F344/N rats and B6C3F1 mice (inhalation studies). Research Triangle Park (NC): U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health. National Toxicology Program, NTP TR 421, NIH Publication No. 93-3152.

Oberdorster G. 1995. The NTP talc inhalation study: a critical appraisal focussed on lung particle overload. Regul Toxicol Pharmacol. 21(2):233-241.

[OECD] Organisation for Economic Co-operation and Development Screening Information Dataset (SIDS). 2004. Synthetic Amorphous Silica and Silicates. SIDS Initial Assessment Report for SIAM 19 [PDF]. Berlin (DE): UNEP Publications. [accessed 2018 Sept].

[OSHA] Occupational Safety and Health Administration. 1999. Talc (not containing asbestos). Chemical Sampling Information. Washington (DC): Occupational Safety and Health Administration (US). [accessed 2017 Nov 7].

Patarino F, Norbedo S, Barbi E, Poli F, Furlan S, Savron F. 2010. Acute Respiratory Failure in a Child after Talc Inhalation. Respiration. 79:340.

Penninkilampi R, Eslick GD. 2018. Perineal talc use and ovarian cancer: A systemic review and meta-analysis. Epidemiology. 29(1):41-49.

Phillips JC, Young PJ, Hardy K, Gangolli SC.1978. Studies on the absorption and disposition of 3H-labelled talc in the rat, mouse, guinea-pig and rabbit. Food Cosmet Toxicol.16(2):161-163.

Pickrell JA, Snipes MB, Benson JM, Hanson RL, Jones RK, Carpenter RL, Thomspon JJ, Hobbs CH, Brown SC. 1989. Talc deposition and effects after 20 days of repeated inhalation exposure of rats and mice to talc. Environ Res. 49:233-245.

Ramlet AA. 1991. A rare complication of ambulatory phlebectomy. Talc Granuloma (French). Phlébologie 44:865-871.

Rasmussen CB, Kjaer SK, Albieri V, Bandera EV, Doherty JA, Høgdall E, Webb PM, Jordan SJ, Rossing MA, Wicklund KG, Goodman MT, Modugno F, Moysich KB, Ness RB, Edwards RP, Schildkraut JM, Berchuck A, Olson SH, Kiemeney LA, Massuger LF, Narod SA, Phelan CM, Anton-Culver H, Ziogas A, Wu AH, Pearce CL, Risch HA, Jensen A; on behalf of the Ovarian Cancer Association Consortium. 2017. Pelvic inflammatory disease and the risk of ovarian cancer and borderline ovarian tumors: a pooled analysis of 13 case-control studies. Am J Epidemiol. 185(1):8-20.

Rasmussen P. 2017. Preliminary talc exposure results. Dec 29, 2017. Unpublished Report Ottawa (ON): Exposure and Biomonitoring Division, Health Canada.

Rasmussen P. 2018. Respirable (PM4) particle concentrations in air while using cosmetics containing talc in Canada, First draft Data Report. July 25, 2018. Unpublished report. Ottawa (ON): Exposure and Biomonitoring Division, Health Canada.

Rosenblatt KA, Szklo M, Rosenshein NB. 1992. Mineral fiber exposure and the development of ovarian cancer. Gynecol Oncol. 45(1):20-25.

Rosenblatt KA, Weiss NS, Cushing-Haugen KL, Wicklund KG, Rossing MA. 2011. Genital powder exposure and the risk of epithelial ovarian cancer. Cancer Causes Control. 22(5):737-742.

Russell RS, Merz RD, Sherman WT, Sivertson JN. 1979. The determination of respirable particles in talcum powder. Cosmet Tox. 17(2):117-122.

Schildkraut JM, Abbott SE, Alberg AJ, Bandera EV, Barnholtz-Sloan JS, Bondy ML, Cote ML, Funkhouser E, Peres LC, Peters ES, et al. 2016. Association between Body Powder Use and Ovarian Cancer: The African American Cancer Epidemiology Study (AACES). Cancer Epidemiol Biomarkers Prev. 25(10):1411-1417.

SDS Search Tool [database]. 2016. Ottawa (ON): Government of Canada. [updated 2016 Sept 15; accessed 2017 Nov 22]. [restricted access].

Shakoor A, Rahatullah A, Shah AA, Zubairi ABS. 2011. Pulmonary talcosis 10 years after brief teenage exposure to cosmetic talcum powder. BMJ Publishing Group. BMJ Case Reports. 2011:bcr0820114597.

Simsek F, Turkeri L, Ilker Y, Kullu S, Akdas A. 1992. Severe obstruction of the urinary tract due to talcum powder granuloma after surgery. A case report. Int Urol Nephrol. 24:31-34.

Statistics Canada. 2016. Data Tables, 2016 Census. Census family structure including stepfamily status (9) and number and age combinations of children (29) for census families with children in private households of Canada, Provinces and Territories, census metropolitan areas and census agglomerations, 2016 and 2100 censuses – 100% data. Ottawa (ON): Government of Canada. [accessed 2017 Nov 23].

Taher MK, Farhat N, Karyakina N, Shilnikova N, Ramoju S, Gravel CA, Krishnan K, Mattison D, Krewski D. 2018. Systematic review of the association between perineal use of talc and ovarian cancer risk. [in preparation].

Terry KL, Karageorgi S, Shvetsov YB, Merritt MA, Lurie G, Thompson PJ, Carney ME, Weber RP, Akushevich L, Lo-Ciganic WH, et al. 2013. Genital powder use and risk of ovarian cancer: a pooled analysis of 8,525 cases and 9,859 controls. Cancer Prev Res. 6(8):811-821.

Trabert B, Pinto L, Hartge P, Kemp T, Black A, Sherman ME, Brinton LA, Pfeiffer RM, Shields MS, Chaturvedi AK, Hildesheim A, and Wentzensen N. 2014. Pre-diagnostic serum levels if inflammation markers and risk of ovarian cancer in the Prostate, Lung, Colorectal and Ovarian Cancer (PLCO) Screening Trial. Gynecol Oncol. 135(2):297-304.

Tukiainen P, Nickels J, Taskinen E, Nyberg M. 1984. Pulmonary granulomatous reaction: talc pneumoconiosis or chronic sarcoidosis? Bri J Ind Med. 41:84-87.

Tzonou A, Polychronopoulou A, Hsieh CC, Rebelakos A, Karakatsani A, Trichopoulos D. 1993. Hair dyes, analgesics, tranquilizers and perineal talc application as risk factors for ovarian cancer. Int J Cancer. 55(3):408-410.

United States. 2016. Federal Register. Banned Devices; Powdered Surgeon's Gloves, Powdered Patient Examination Gloves, and Absorbable Powder for Lubricating a Surgeon's Glove, A Rule by the Food and Drug Administration on 12/19/2016. US: Federal Register (US). Vol. 81, No. 243. 21 CFR 878. p. 91722-91731 [accessed 2018 Jan 3].

[U.S. EPA] United States Environmental Protection Agency. 1992. Health Assessment Document for Talc. Washington (D.C.): Office of Research and Development. Report No. EPA 600/8-91/217.

[U.S. EPA] United States Environmental Protection Agency. 1994. Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry. Research Triangle Park (NC): U.S. EPA, Environmental Criteria and Assessment Office, Office of Health and Environmental Assessment, Office of Research and Development.

[U.S. EPA] United States Environmental Protection Agency. 2005. Guidelines for Carcinogen Risk Assessment [PDF]. Washington (D.C.): U.S. EPA, EPA/630/P-03/001F.

[U.S. EPA] United States Environmental Protection Agency. 2009. Risk Assessment Guidance for Superfund Volume I: Human Health Evaluation Manual (Part F, Supplemental Guidance for Inhalation Risk Assessment). Washington (D.C.): U.S. EPA, Office of Superfund Remediation and Technology Innovation.

[U.S. EPA] United States Environmental Protection Agency. 2011. Exposure Factors Handbook 2011 Edition (Final Report). Washington (D.C.): U.S. EPA, EPA/600/R-09/052F.

[U.S. FDA] United States Food and Drug Administration. 2015. Select Committee on GRAS Substances (SCOGS) Opinion: Silicates. Silver Spring (MD): U.S. Food and Drug Administration. [accessed 2018 Aug 17]

[U.S. FDA] United States Food and Drug Administration. 2016. About the GRAS Notification Program. Silver Spring (MD): US Food and Drug Administration. [accessed 2017 Mar 12].

[USGS] United States Geological Survey. 2000. U.S. Talc-Baby Powder and Much More [PDF]. Reston (VA): US Geological Survey. USGS Fact Sheet FS-065-00. [accessed 2017 May 29].

[USGS] United States Geological Survey. 2018. Mineral Commodity Summaries. Talc and Pyrophyllite [PDF]. Reston (VA): US Geological Survey. [accessed 2018 August 13].

[USP] US Pharmacopeia. 2011. USP Monographs: Talc. Talc Revision Bulletin Official August 1, 2011 [PDF]. US: The United States Pharmacopeial Convention. [accessed 2018 May 3].

Vallyathan NV, Craighead JE.1981. Pulmonary pathology in workers exposed to nonasbestiform talc. Hum Pathol.12(1):28-35.

Vanderhyden BC, Shaw TJ, Ethier JF. 2003. Animal models of ovarian cancer. Reprod Biol Endocrinol. 1:67.

Venter PF, Iturralde M. 1979. Migration of a particulate radioactive tracer from the vagina to the peritoneal cavity and ovaries. S Afr Med J. 55(23):917-919.

Wadaan MAM. 2009. Effects of repeated exposure to talcum powder on rabbit skin. Indian J Appl Pure Biol. 24(1):111-115.

Wagner JC, Berry G, Cooke TJ, Hill RJ, Pooley FD, Skidmore JW. 1977. Animal experiments with talc. Inhaled Particles. 4 Pt 2:647-654.

Warheit, DB, Kreiling R, Levy LS. 2016. Relevance of the rat lung tumor response to particle overload for human risk assessment-Update and interpretation of new data since ILSI 2000. Toxicology. 374:42-59.

Wehner AP, Tanner TM, Buschbom RL. 1977a. Absorption of ingested talc by hamsters. Food Cosmet Toxicol.15(5):453-455.

Wehner AP, Wilkerson CL, Cannon WC, Buschbom RL, Tanner TM. 1977b. Pulmonary deposition, translocation and clearance of inhaled neutron-activated talc in hamsters. Food Cosmet Toxicol.15(5):213-224.

Wehner AP, Hall AS, Weller RE, Lepel EA, Schirmer RE. 1985. Do particles translocate from the vagina to the oviducts and beyond? Food Chem Toxicol. 23(3):367-372.

Wehner AP, Weller RE, Lepel EA. 1986. On talc translocation from the vagina to the oviducts and beyond. Food Chem Toxicol. 24(4):329-338.

Wehner AP. 2002. Cosmetic talc should not be listed as a carcinogen: comments on NTP's deliberations to list talc as a carcinogen. Regul Toxicol Pharmacol. 36:40-50.

Whittemore AS, Wu ML, Paffenbarger RS Jr, Sarles DL, Kampert JB, Grosser S, Jung DL, Ballon S, Hendrickson M. 1988. Personal and environmental characteristics related to epithelial ovarian cancer. I. Exposures to talcum powder, tobacco, alcohol, and coffee. Am J Epidemiol. 128(6):1228-1240.

[WHO, UNFPA, FHI] World Health Organization, United Nations Population Fund, Family Health International. 2013. Male latex condom. Specification, prequalitfication and guidelines for procurement, 2010, revised April 2013. Geneva (CH): World Health Organization. [accessed 2017 Dec 20].

Wild P, Leodolter K, Refregier M, Schmidt H, and Bourgkard E. 2008. Effect of talc dust on respiratory health: results of a longitudinal survey of 378 French and Austrian talc workers. Occup Environ Med. 65: 261-267.

Wong C, Hempling RE, Piver MS, Natarajan N, Mettlin CJ. 1999. Perineal talc exposure and subsequent epithelial ovarian cancer: a case-control study. Obstet Gynecol. 93(3):372-376.

Wu AH, Pearce CL, Tseng CC, Templeman C, Pike MC. 2009. Markers of inflammation and risk of ovarian cancer in Los Angeles County. Int J Cancer. 124(6):1409-1415.

Wu AH, Pearce CL, Tseng CC, Pike MC. 2015. African Americans and Hispanics Remain at Lower Risk of Ovarian Cancer Than Non-Hispanic Whites after Considering Nongenetic Risk Factors and Oophorectomy Rates. Cancer Epidemiol Biomarkers Prev. 24(7):1094-1100.

Zazenksi R, Ashton WH, Briggs D, Chudkowski M, Kelse JW, MacEachern L, McCarthy EF, Norhauser MA, Roddy MT, Teetsel NM, Wells AB, Gettings SD. 1995. Talc: Occurrence, Characterization, and Consumer Applications. Reg Pharm Tox. 21:218-229.

Appendix A. Inhalation exposure estimates

Table A-1. Estimated inhalation exposure concentrations from self-care products containing loose powder talc available to consumers

Scenario	Talc product conc.[a]	Study[b] conc. (mg/m^3)	CA[b] (mg/m^3)	ET[c] (hr/d)	EF[d] (d/yr)	ED[e] (yr)	EC adjusted (mg/m^3)[b]
Baby powder, infants	100 %	1.36	1.36	0.125	365	4	0.0071
Baby powder, adults	100 %	1.36	1.36	0.125	365	8	0.0071
Body powder, adults	100 %	1.36	1.36	0.083	365	58	0.0047
Face powder, adults	100 %	1.36	1.36	0.083	365	58	0.0047
Foot powder, adults	97 %	1.36	1.32	0.083	274	58	0.0034
Dry hair shampoo, adults	100 %	1.36	1.36	0.083	84	58	0.0011

Abbreviations: Conc., concentration; CA, concentration in air per event; ET, exposure time; EF, exposure frequency; ED, exposure duration; EC, adjusted exposure concentration.

[a] Highest concentration of talc found per product type from notifications submitted under the *Cosmetic Regulations* to Health Canada for talc, DPD [modified 2018], email from the Therapeutic Products Directorate, Health Canada, to the Existing Substances Risk Assessment Bureau, Health Canada, dated March 20, 2017, unreferenced; LNHPD [modified 2018], email from the Non-prescription and Natural Health Products Directorate, Health Canada, to the Existing Substances Risk Assessment Bureau, Health Canada, dated March 20, 2017, unreferenced; Fiume et al. 2015; Household Product Database 1993-; CPCat 2014; CPID 2017; SDS Search Tool 2016.

[b] Average by subject from Anderson et al. 2107 and Rasmussen 2018 (unpublished). CA = average study concentration × maximum talc concentration in product.

[c] ET is 5 minutes/application based on median time spent in the bathroom following a shower or bath (U.S. EPA 2011) × number of applications/day, whereby baby powder assumes 1.5 applications/day (CTFA 1983); the rest assume 1 application/day.

[d] EF is assumed to be daily for baby, body (U.S. EPA 2011) and face powder (Ficheux et al. 2015); foot powder 0.75 times/day or 274 times/year (Ficheux et al. 2015); dry hair shampoo 0.23 times/day or 84 times/year (Ficheux et al. 2015).

[e] Assumed infant wears diapers up to 4 years, adult exposure to baby powder from diapering children, 4 years per child and assume 2 children per family (Statistics Canada 2016), adult exposure for body powder, and foot powder (80 years lifetime, 12 years child).

[f] Adjusted exposure concentration is calculated as per Equation 8 in the U.S. EPA 2009 guidance document "Risk Assessment Guidance for Superfund Volume I: Human Health Evaluation Manual," where EC = (CA × ET × EF × ED)/AT, and AT = averaging time, which is on the basis of ED × 365 days/year × 24 hours/day.